

GenCore version 5.1.4.P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 30, 2003, 14:34:49 ; Search time 16 Seconds
(without alignments)
1197.228 Million cell updates/sec

Title: US-09-513-999c-3792_COPY_51_161

Perfect score: 208
Sequence: 1 atggctgagctcttttgcctt.....gcctgagtgctgcttact 111

Scoring table:

| | |
|----------|---------------------------|
| BLOSUM62 | Xgapop 10.0 , Xgapext 0.5 |
| | Ygapop 10.0 , Ygapext 0.5 |
| | Fgapop 6.0 , Fgapext 7.0 |
| | Delpop 6.0 , Delpext 7.0 |

Searched: 328255 segs, 86286685 residues

Total number of hits satisfying chosen parameters: 656510

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+np.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09513999/runat_30042003_143404_25670/app-query.fasta.1.263
-DB=Published Applications_AA -QMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -SPART=1 -END=-1 -MATRIX=Dlosum62
-TRANS-human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=ext -THR_MAX=100
-THR_MIN=0 -ALLEN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09513999_@CGN_1_1_17_@runat_30042003_143404_25670
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMMP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCOT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCOTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| 1 | 62 | 29.8 | 4842 | 9 | US-10-184-644-289 |
| 2 | 62 | 29.8 | 4842 | 9 | US-10-184-634-289 |
| 3 | 60.5 | 29.1 | 77 | 9 | US-09-965-528-1 |
| 4 | 60 | 28.8 | 695 | 9 | US-10-184-644-567 |

ALIGNMENTS

RESULT 1
US-10-184-644-289
; Sequence 289, Application US/10184644
; Publication No. US20030044930A1
GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 289
; LENGTH: 4842
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-289

| 5 | 60 | 28.8 | 695 | 9 | US-10-184-634-567 | Sequence 567, App |
|----|------|------|------|--|---------------------|--------------------|
| 6 | 59 | 28.4 | 1768 | 9 <td>US-10-184-644-343</td> <td>Sequence 343, App</td> | US-10-184-644-343 | Sequence 343, App |
| 7 | 59 | 28.4 | 1768 | 9 <td>US-10-184-634-343</td> <td>Sequence 343, App</td> | US-10-184-634-343 | Sequence 343, App |
| 8 | 59 | 28.4 | 1941 | 9 <td>US-10-123-155-165</td> <td>Sequence 165, App</td> | US-10-123-155-165 | Sequence 165, App |
| 9 | 59 | 28.4 | 3162 | 9 <td>US-10-123-155-111</td> <td>Sequence 111, App</td> | US-10-123-155-111 | Sequence 111, App |
| 10 | 59 | 28.4 | 3192 | 9 <td>US-10-123-155-75</td> <td>Sequence 75, App</td> | US-10-123-155-75 | Sequence 75, App |
| 11 | 58.5 | 29.0 | 126 | 9 <td>US-09-738-626-6097</td> <td>Sequence 6097, App</td> | US-09-738-626-6097 | Sequence 6097, App |
| 12 | 58 | 27.9 | 2334 | 9 <td>US-10-123-155-129</td> <td>Sequence 129, App</td> | US-10-123-155-129 | Sequence 129, App |
| 13 | 58 | 27.9 | 2336 | 9 <td>US-10-123-155-383</td> <td>Sequence 383, App</td> | US-10-123-155-383 | Sequence 383, App |
| 14 | 57 | 28.2 | 990 | 10 <td>US-09-823-356-9</td> <td>Sequence 9, App</td> | US-09-823-356-9 | Sequence 9, App |
| 15 | 57 | 27.4 | 2476 | 9 <td>US-10-184-644-585</td> <td>Sequence 585, App</td> | US-10-184-644-585 | Sequence 585, App |
| 16 | 57 | 27.4 | 2476 | 9 <td>US-10-184-634-585</td> <td>Sequence 585, App</td> | US-10-184-634-585 | Sequence 585, App |
| 17 | 57 | 27.4 | 3951 | 9 <td>US-10-184-634-119</td> <td>Sequence 119, App</td> | US-10-184-634-119 | Sequence 119, App |
| 18 | 57 | 27.4 | 3951 | 9 <td>US-10-184-634-119</td> <td>Sequence 119, App</td> | US-10-184-634-119 | Sequence 119, App |
| 19 | 56.5 | 27.2 | 61 | 9 <td>US-09-809-391-398</td> <td>Sequence 398, App</td> | US-09-809-391-398 | Sequence 398, App |
| 20 | 56.5 | 27.2 | 1679 | 9 <td>US-10-123-155-375</td> <td>Sequence 375, App</td> | US-10-123-155-375 | Sequence 375, App |
| 21 | 56 | 27.7 | 56 | 9 <td>US-09-796-692-748</td> <td>Sequence 748, App</td> | US-09-796-692-748 | Sequence 748, App |
| 22 | 56 | 27.7 | 56 | 9 <td>US-10-040-862-748</td> <td>Sequence 748, App</td> | US-10-040-862-748 | Sequence 748, App |
| 23 | 56 | 26.9 | 80 | 10 <td>US-09-864-761-43395</td> <td>Sequence 43395, A</td> | US-09-864-761-43395 | Sequence 43395, A |
| 24 | 56 | 26.9 | 190 | 10 <td>US-10-125-540-305</td> <td>Sequence 305, App</td> | US-10-125-540-305 | Sequence 305, App |
| 25 | 56 | 26.9 | 190 | 10 <td>US-09-764-870-305</td> <td>Sequence 305, App</td> | US-09-764-870-305 | Sequence 305, App |
| 26 | 56 | 26.9 | 396 | 10 <td>US-09-953-956-13</td> <td>Sequence 13, App</td> | US-09-953-956-13 | Sequence 13, App |
| 27 | 56 | 26.9 | 396 | 10 <td>US-09-953-956-13</td> <td>Sequence 13, App</td> | US-09-953-956-13 | Sequence 13, App |
| 28 | 56 | 26.9 | 396 | 12 <td>US-10-114-464-13</td> <td>Sequence 13, App</td> | US-10-114-464-13 | Sequence 13, App |
| 29 | 56 | 26.9 | 1666 | 9 <td>US-10-073-912-12</td> <td>Sequence 12, App</td> | US-10-073-912-12 | Sequence 12, App |
| 30 | 56 | 26.9 | 2148 | 9 <td>US-10-184-634-507</td> <td>Sequence 507, App</td> | US-10-184-634-507 | Sequence 507, App |
| 31 | 56 | 26.9 | 2148 | 9 <td>US-10-184-634-507</td> <td>Sequence 507, App</td> | US-10-184-634-507 | Sequence 507, App |
| 32 | 56 | 26.9 | 3322 | 9 <td>US-10-184-634-489</td> <td>Sequence 489, App</td> | US-10-184-634-489 | Sequence 489, App |
| 33 | 56 | 26.9 | 3322 | 9 <td>US-10-184-634-489</td> <td>Sequence 489, App</td> | US-10-184-634-489 | Sequence 489, App |
| 34 | 56 | 26.9 | 1836 | 9 <td>US-10-073-912-17</td> <td>Sequence 17, App</td> | US-10-073-912-17 | Sequence 17, App |
| 35 | 55.5 | 26.7 | 147 | 9 <td>US-10-042-296-6</td> <td>Sequence 6, App</td> | US-10-042-296-6 | Sequence 6, App |
| 36 | 55.5 | 26.7 | 1572 | 9 <td>US-10-184-634-65</td> <td>Sequence 65, App</td> | US-10-184-634-65 | Sequence 65, App |
| 37 | 55.5 | 26.7 | 1572 | 9 <td>US-10-184-634-65</td> <td>Sequence 65, App</td> | US-10-184-634-65 | Sequence 65, App |
| 38 | 55.5 | 26.7 | 2037 | 9 <td>US-10-184-634-13</td> <td>Sequence 13, App</td> | US-10-184-634-13 | Sequence 13, App |
| 39 | 55.5 | 26.7 | 2037 | 9 <td>US-10-184-634-13</td> <td>Sequence 13, App</td> | US-10-184-634-13 | Sequence 13, App |
| 40 | 55 | 27.2 | 586 | 9 <td>US-09-764-686-624</td> <td>Sequence 624, App</td> | US-09-764-686-624 | Sequence 624, App |
| 41 | 55 | 27.2 | 586 | 9 <td>US-09-955-999-64</td> <td>Sequence 64, App</td> | US-09-955-999-64 | Sequence 64, App |
| 42 | 55 | 26.4 | 1738 | 9 <td>US-10-184-634-239</td> <td>Sequence 239, App</td> | US-10-184-634-239 | Sequence 239, App |
| 43 | 55 | 26.4 | 1738 | 9 <td>US-10-184-634-239</td> <td>Sequence 239, App</td> | US-10-184-634-239 | Sequence 239, App |
| 44 | 55 | 26.4 | 1968 | 9 <td>US-10-123-155-163</td> <td>Sequence 163, App</td> | US-10-123-155-163 | Sequence 163, App |
| 45 | 55 | 26.4 | 2806 | 9 <td>US-10-123-155-201</td> <td>Sequence 201, App</td> | US-10-123-155-201 | Sequence 201, App |

1 APPLICANT: Baker, Kevin F.
 2 APPLICANT: Chen, Jian
 3 APPLICANT: Desnoyers, Luc
 4 APPLICANT: Goddard, Audrey
 5 APPLICANT: Godowski, Paul J.
 6 APPLICANT: Gurney, Austin L.
 7 APPLICANT: Pan, James
 8 APPLICANT: Smith, Victoria
 9 APPLICANT: Watanabe, Colin K.
 10 APPLICANT: Wood, William I.
 11 APPLICANT: Zhang, Zemin
 12 TITLE OF INVENTION: AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 13 FILE REFERENCE: P3430R1C227
 14 CURRENT APPLICATION NUMBER: US/10/184,644
 15 PRIOR APPLICATION removed - See File Wrapper or Palm
 16 NUMBER OF SEQ ID NOS: 612
 17 SEQ ID NO 567

```
; LENGTH: 695
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-567

Alignment Scores:
Pred. No.: 20.9      Length: 695
Score: 60.00      Matches: 13
Percent Similarity: 48.39%      Conservative: 2
Best Local Similarity: 41.94%      Mismatches: 15
Query Match: 28.85%      Indels: 1
DB: 9      Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-184-644-567 (1-695)

OY 15 TGCCCTGAGAGATCTTTTCATCTTTCAGAGACTTCGGCGCGAGATGTAACACT 74
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 606 CysCysAlaGlyGlyAlaCysCysAlaAlaGlyAlaAlaGlyThrGlyAlaCysGly 625
OY 75 CCTGGCTCTGTGTGTGCTGAGTGGCTGC 105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 ThrGlyGly-CysThrCysCysAlaGlyCys 635

RESULT 5
; Sequence 567, Application US/10184634
; Publication No. US2003006864A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 567
; LENGTH: 695
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-567

Alignment Scores:
Pred. No.: 20.9      Length: 695
Score: 60.00      Matches: 13
Percent Similarity: 48.39%      Conservative: 2
Best Local Similarity: 41.94%      Mismatches: 15
Query Match: 28.85%      Indels: 1
DB: 9      Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-184-634-567 (1-695)

OY 15 TGCCCTGAGAGATCTTTTCATCTTTCAGAGACTTCGGCGCGAGATGTAACACT 74
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 606 CysCysAlaGlyGlyAlaCysCysAlaAlaGlyAlaAlaGlyThrGlyAlaCysGly 625
OY 75 CCTGGCTCTGTGTGTGCTGAGTGGCTGC 105
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 ThrGlyGly-CysThrCysCysAlaGlyCys 635

RESULT 6
US-10-184-644-343
; Sequence 343, Application US/10184644
```

```
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 343
; LENGTH: 1768
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-343

Alignment Scores:
Pred. No.: 30.9      Length: 1768
Score: 59.00      Matches: 13
Percent Similarity: 50.00%      Conservative: 3
Best Local Similarity: 40.62%      Mismatches: 9
Query Match: 28.37%      Indels: 7
DB: 9      Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-184-644-343 (1-1768)

OY 15 TGCCCTGAGAGATCTTTTCATCTTTCAGAGACTTCGGCGCGAGATGTAACACT 74
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 CysCysAlaGly-----GlyAlaCysGlyGlyGlyCysGlyThr 572
OY 75 CCTGGCTCTGTGTGTGCTGAGTGGCTGCTCT 108
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 573 GlyGlyAla-CysAlaCysCysThrGlyCysThr 583

RESULT 7
US-10-184-634-343
; Sequence 343, Application US/10184634
; Publication No. US2003006864A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 343
; LENGTH: 1768
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-343
```

```
Alignment Scores: 30.9 Length: 1768
Pred. No.: 59.00 Matches: 13
Score: 59.00% Conservative: 3
Percent Similarity: 50.00% Mismatches: 9
Best Local Similarity: 40.62% Indels: 7
Query Match: 28.37% Gaps: 1
DB: 9

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-184-634-343 (1-1768)
QY 15 TGCGCTGCGAGATCTTTTCATCTTTCAGGAGCTTCGCGCGAGATATGAAACT 74
DB 559 CysGSAIAGly-----GlyAlaCysGlyGlyGlyCysGysGlyThr 572
QY 75 CCGGCTCTGTGTGTGCGCTGAGTGCCTCTCT 108
DB 573 GlyGlyAla-CysAlaCysCysThrGlyCysThr 583

RESULT 8
US-10-123-155-165
Sequence 165, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3330R1C30
CURRENT APPLICATION NUMBER: US/10/123,155
PRIORITY FILING DATE: 2002-04-15
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 550
SEQ ID NO 165
LENGTH: 1941
TYPE: DNA
ORGANISM: Homo Sapien
US-10-123-155-165

Alignment Scores: 31.2 Length: 1941
Pred. No.: 59.00 Matches: 11
Score: 60.87% Conservative: 3
Percent Similarity: 47.83% Mismatches: 8
Best Local Similarity: 28.37% Indels: 1
Query Match: 28.37% Gaps: 0
DB: 9

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-123-155-165 (1-1941)
QY 45 GGGAGCTTGGGGCGGCGGATGTAAGACCTGCTCTGTGTGCTGCTGAGTGGCTG 104
DB 1815 GlyGlyThrGlyGlyGlyCysCysCysThrAlaGlyGlyCysThrGlyGlyCys 1834
QY 105 CTCTACT 111
DB 1834 sThrThr 1836

RESULT 9
US-10-123-155-111
```

```
; Sequence 111, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin

US-10-123-155-75
Sequence 75, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-123-155-111 (1-3162)
QY 3 GGGTGATCTTTTGGCTTCAGAGATCTTTTCATCTTTCAGGAGCTTCGCGCGGA 62
DB 1320 GlyThrCysCysCysAlaGlyAlaGlyCysCysThrGlyGlyGlyAlaGly 1339
QY 63 GATGTAAACCTCGGCTCTCTGTGTGCTGCTGAGTGCCTGCTCTACT 111
DB 1340 CysCysGlyThrAlaGlyGlyCys-CysCysAlaAlaCysCysThr 1355

RESULT 10
US-10-123-155-75
Sequence 75, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Beresini, Maureen
APPLICANT: Deforge, Laura
APPLICANT: Desnoyers, Luc
APPLICANT: Filvaroff, Ellen
APPLICANT: Gao, Wei-Qiang
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Sherwood, Steven
APPLICANT: Smith, Victoria
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K
APPLICANT: Wood, William
APPLICANT: Zhang, Zemin

Alignment Scores: 32.8 Length: 3162
Pred. No.: 59.00 Matches: 15
Score: 43.24% Conservative: 1
Percent Similarity: 40.54% Mismatches: 20
Best Local Similarity: 28.37% Indels: 1
Query Match: 28.37% Gaps: 0
DB: 9
```

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123.155
; PRIOR FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 75
; LENGTH: 3192
; TYPE: DNA
; ORGANISM: Homo Sapien
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1428,1431
; OTHER INFORMATION: unknown base
; US-10-123-155-75

Alignment Scores:
Pred. No.: 32.9 Length: 3192
Score: 59.00 Matches: 13
Percent Similarity: 53.12% Conservative: 4
Best Local Similarity: 40.62% Mismatches: 14
Query Match: 28.37% Indels: 1
DB: 9 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-123-155-75 (1-3192)
QY 15 TGGCTTCAGAGATCTTTTCATCTTTCGAGGAGCTTCGGCGGAGTATGTAACACT 74
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1554 CysCysAlaGlaGlyAlaGlyAlaAlaCysAlaGlyCysAlaGlyAlaGlyThrcysAlaGly 1573
QY 75 CCGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 108
    ::||| ||| ||| ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1574 ThrThrAla-CysAlaCysCysThrGlyCysAla 1584

RESULT 11
US-09-738-626-6097
; Sequence 6097, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIICO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6097
; LENGTH: 126
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-6097

Alignment Scores:
Pred. No.: 27.1 Length: 126
Score: 58.50 Matches: 12
Percent Similarity: 59.26% Conservative: 4
Best Local Similarity: 44.44% Mismatches: 8
```

```

Query Match: 28.96% Indels: 3
DB: 9 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-738-626-6097 (1-126)
QY 91 ACACACAGAGACCCGAGGTTTTCATATCTCCGCGCCAGAGTCCCT-----GCA 41
    ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 90 SerHisHisLysAspGlyLeuThrArgLeuProAlaProIuLeuProAsnAlaIaIle 109
QY 40 AAGATGAAAAAGAAVCTGCA 20
    ||| ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110 LysAlaArgLysAsnProAla 116

RESULT 12
US-10-123-155-129
; Sequence 129, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tunas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C30
; CURRENT APPLICATION NUMBER: US/10/123.155
; PRIOR FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 129
; LENGTH: 2334
; TYPE: DNA
; ORGANISM: Homo Sapien
; US-10-123-155-129

Alignment Scores:
Pred. No.: 42.6 Length: 2334
Score: 58.00 Matches: 13
Percent Similarity: 58.33% Conservative: 1
Best Local Similarity: 54.17% Mismatches: 7
Query Match: 27.88% Indels: 3
DB: 9 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x US-10-123-155-129 (1-2334)
QY 42 GCAGGAGACTTCGGGGCGGAGATATGTAACACTCCTGGGCTCTGTGTGTCCTGATG 101
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 277 AlaGlyThrGlyAlaGlyCysCys-----ThrGlyAla-CysAlaCysThrGlyG 294
QY 102 CCGCTCTACT 111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 294 YCysGlyThr 297

RESULT 13
US-10-123-155-383
; Sequence 383, Application US/10123155
; Publication No. US20030068794A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
```

```

; SEQ ID NO 9
; LENGTH: 950
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20010025098A1 1794154
US-09-823-356-9

Alignment Scores:
Pred. No.:          52           Length:      950
Score:              57.00        Matches:     15
Percent Similarity: 58.62%       Conservative: 2
Best Local Similarity: 51.72%    Mismatches:   8
Query Match:        28.22%       Indels:       4
DB:                 10           Gaps:         2

US-09-513-999C-3792_COPY_51_161 (1-111) x US-09-823-356-9 (1-950)
QY      88 CACAGAGCCGACGAGGTTCATCACTCCGGCCCCACAGAAGTCCCTCGAAAGATGA--- 32
Db      475 HtSagAspproAShValVal-----ProlypProProlypProAlatysGlutylsPro 492
QY      31 ---AMGAATCCTGCAGAGCAAGAATC 8
Db      493 ProlysLysLysAlaGlnAspLysIle 501

RESULT 15
US-10-184-644-585
; Sequence 585, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zhenli
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 585
; LENGTH: 2476
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-585

Alignment Scores:
Pred. No.:          57.5           Length:     2476
Score:              57.00        Matches:     13
Percent Similarity: 50.00%       Conservative: 2
Best Local Similarity: 43.33%    Mismatches:   8
Query Match:        27.40%       Indels:       7
DB:                 9           Gaps:         1

US-09-513-999C-3792_COPY_51_161 (1-111) x US-10-184-644-585 (1-2476)
QY      42 GCAGGAGCTCTGCGGGCGGAGATGTAATAACTCCGCG-----TCT 83
Db      60 ALaclgThrlglgylgylglaAcysalaThrThrglCysThrlglalAcysalaala 79
QY      84 CTGTGTGTGCCCTGAGTGCGTCTCTACT 111
Db      80 ThrcVcs-CvscVsglVglvcvsthrThr 88

```

Wed Apr 30 15:10:00 2003

us-09-513-999C-3792_copy_51_161.n2p.raph

Page 7

Search completed: April 30, 2003, 14:37:30
Job time : 21 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 30, 2003, 14:35:15 ; Search time 12 Seconds

(without alignments)
544.324 Million cell updates/sec

Title: US-09-513-999c-3792_COPY_51_161

Perfect score: 208
Sequence: 1 atgagtgatcttttgcctt.....gccctgagtgctgtctact 111

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+np.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US0513999/rnat_30042003_143402_25630/app_query.fasta_1.263
-DB=Issued Patents_AA -QFMT=fastin -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptio -NORMEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0513999_ECGN_1_1_28_@rnat_30042003_143402_25630 -NCPD=6 -ICPD=3
-NO_XLPEXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMCOUT=120
-NARN_TIMCOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents_AA:*
1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------------|
| c 1 | 63.5 | 31.4 | 640 | 4 | US-09-262-773-4 |
| c 2 | 63.5 | 31.4 | 648 | 4 | US-09-262-773-2 |
| c 3 | 60.5 | 29.1 | 218 | 2 | US-08-399-889-25 |
| c 4 | 60.5 | 29.1 | 218 | 3 | US-09-167-364-25 |
| c 5 | 60.5 | 29.1 | 218 | 4 | US-09-439-897-4 |
| c 6 | 60.5 | 29.1 | 268 | 4 | US-09-589-927-6 |
| c 7 | 60.5 | 29.1 | 268 | 4 | US-09-277-665-6 |
| c 8 | 60.5 | 29.1 | 471 | 2 | US-08-399-889-24 |
| c 9 | 60.5 | 29.1 | 471 | 3 | US-09-167-364-24 |
| c 10 | 60.5 | 29.1 | 471 | 4 | US-09-439-897-2 |
| c 11 | 59 | 29.2 | 1297 | 2 | US-08-290-731C-4 |
| c 12 | 56.5 | 27.2 | 61 | 4 | US-09-149-476-398 |

| | | | | | | |
|------|------|------|------|---|-------------------|-------------------|
| 13 | 56 | 26.9 | 349 | 4 | US-09-032-523-3 | Sequence 3, Appl |
| 14 | 56 | 26.9 | 396 | 1 | US-08-208-007A-13 | Sequence 13, Appl |
| 15 | 56 | 26.9 | 396 | 4 | US-09-032-523-9 | Sequence 9, Appl |
| 16 | 56 | 26.9 | 396 | 4 | US-08-915-095A-13 | Sequence 13, Appl |
| 17 | 56 | 26.9 | 396 | 4 | US-08-798-096-13 | Sequence 13, Appl |
| 18 | 56 | 26.9 | 396 | 4 | US-08-798-095A-13 | Sequence 13, Appl |
| 19 | 54 | 26.0 | 1720 | 2 | US-08-477-451-12 | Sequence 12, Appl |
| c 20 | 52.5 | 26.0 | 976 | 3 | US-08-560-005-2 | Sequence 2, Appl |
| c 21 | 52.5 | 26.0 | 976 | 3 | US-09-195-868-14 | Sequence 14, Appl |
| c 22 | 52.5 | 26.0 | 976 | 4 | US-09-418-540-2 | Sequence 2, Appl |
| c 23 | 52.5 | 26.0 | 1187 | 4 | US-08-664-962B-8 | Sequence 8, Appl |
| c 24 | 52.5 | 26.0 | 1187 | 3 | US-09-311-743-8 | Sequence 8, Appl |
| c 25 | 52.5 | 26.0 | 1189 | 3 | US-09-195-868-15 | Sequence 15, Appl |
| c 26 | 52.5 | 26.0 | 1229 | 3 | US-09-195-868-28 | Sequence 28, Appl |
| c 27 | 52 | 25.0 | 418 | 2 | US-08-494-807-18 | Sequence 18, Appl |
| c 28 | 52 | 25.0 | 418 | 5 | PCT-US96-10986-18 | Sequence 18, Appl |
| c 29 | 52 | 25.0 | 2088 | 4 | US-09-548-372D-13 | Sequence 13, Appl |
| c 30 | 52 | 25.0 | 2088 | 4 | US-09-548-367D-13 | Sequence 13, Appl |
| c 31 | 51.5 | 24.8 | 396 | 2 | US-09-061-337-12 | Sequence 12, Appl |
| c 32 | 51.5 | 24.8 | 396 | 2 | US-09-122-129-12 | Sequence 12, Appl |
| c 33 | 51.5 | 24.8 | 396 | 3 | US-09-340-991-12 | Sequence 12, Appl |
| c 34 | 51.5 | 24.8 | 396 | 4 | US-08-974-609-12 | Sequence 12, Appl |
| c 35 | 51.5 | 24.8 | 396 | 4 | US-09-549-098-12 | Sequence 12, Appl |
| c 36 | 51.5 | 24.8 | 1417 | 4 | US-08-900-230-3 | Sequence 3, Appl |
| c 37 | 51 | 24.5 | 24 | 4 | US-09-443-501A-21 | Sequence 21, Appl |
| c 38 | 51 | 25.2 | 1119 | 4 | US-09-396-651B-2 | Sequence 2, Appl |
| c 39 | 50.5 | 25.0 | 564 | 1 | US-07-872-644-53 | Sequence 53, Appl |
| c 40 | 50.5 | 25.0 | 564 | 1 | US-08-297-494-53 | Sequence 53, Appl |
| c 41 | 50.5 | 25.0 | 564 | 1 | US-08-297-510-53 | Sequence 53, Appl |
| c 42 | 50.5 | 25.0 | 564 | 1 | US-08-479-532-53 | Sequence 53, Appl |
| c 43 | 50.5 | 25.0 | 564 | 1 | US-08-455-526-53 | Sequence 53, Appl |
| c 44 | 50.5 | 25.0 | 564 | 1 | US-08-455-525-53 | Sequence 53, Appl |
| c 45 | 50.5 | 25.0 | 564 | 3 | US-09-139-491-53 | Sequence 53, Appl |

ALIGNMENTS

RESULT 1
US-09-262-773-4
; Sequence 4, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: MYriad 3
; CURRENT APPLICATION NUMBER: US/09/262, 773
; CURRENT FILING DATE: 1998-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 640
; TYPE: PRT
; ORGANISM: human
US-09-262-773-4

Alignment Scores:
Pred. No.: 1.15
Score: 63.50
Percent Similarity: 66.67%
Best Local Similarity: 44.44%
Query Match: 31.44%
DB: 4
Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-262-773-4 (1-640)

QY 110 GAGAGCAGCCCTGAGGCACACAGAGACCCAG-----AGTTTACTACTACGGCG 57
DB 292 IlegInguPro--GlnGluThrInGInuProGluIleLeuSerPheThrTyThrIle 310

```

OY      56 CCCAGAGTCCCTGCAAGATGAAAGAAATCTGCAAGGCAAAAGAT 9
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      311 AsparGser-----LysAspGluGluGluGluGluGluGluGluAsp 324

RESULT 2
US-09-262-773-2
; Sequence 2, Application US/09262773
; Patent No. 6225451
; GENERAL INFORMATION:
; APPLICANT: Ballinger, Dennis G.
; APPLICANT: Ding, Wei
; APPLICANT: Wagner, Susanne
; APPLICANT: Hess, Mark A.
; TITLE OF INVENTION: CHROMOSOME 11-LINKED CORONARY HEART DISEASE
; FILE REFERENCE: Myriad 3
; CURRENT APPLICATION NUMBER: US/09/262,773
; CURRENT FILING DATE: 1999-03-04
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 648
; TYPE: PRT
; ORGANISM: human
US-09-262-773-2

Alignment Scores:
Pred. No.: 1.15      Length: 648
Score: 63.50      Matches: 16
Percent Similarity: 66.67%      Conservative: 8
Best Local Similarity: 44.44%      Mismatches: 7
Query Match: 31.44%      Indels: 5
DB: 4      Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-262-773-2 (1-648)

OY      110 GTACGAGCAGCAGCTGACGACACAGACCCAGG-----AGTTTAACTACTCCGCG 57
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      300 lIeInGluPro--GInGluThGInGluProGluIleuSerPheTrItyrInGly 318

OY      56 CCCAGAGTCCCTGCAAGATGAAAGAAATCTGCAAGGCAAAAGAT 9
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      319 AsparGser-----LysAspGluGluGluGluGluGluGluGluAsp 332

RESULT 3
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-399-889-25

Alignment Scores:
Pred. No.: 2.49      Length: 218
Score: 60.50      Matches: 17
Percent Similarity: 56.76%      Conservative: 4
Best Local Similarity: 45.95%      Mismatches: 11
Query Match: 29.09%      Indels: 5
DB: 2      Gaps: 3

```

```

US-09-513-999c-3792_COPY_51_161 (1-111) x US-08-399-889-25 (1-218)

OY      3 GGGTGATCTTTTGCCCTTCAGAGATCTTTTCATCTTT-----GCAGGACTTCT 53
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 GlyTrpIleSerLeuTrpIleGlyPheSerPheIleMetPheThrSerAlaIleGlySerIu 140

OY      54 GGGGCCGGA---GTATGTAAGACTCTGGGCTCTGTGTGTGCTGAGTGG 101
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      141 GlyAlaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGluGluPhe 156

RESULT 4
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

Alignment Scores:
Pred. No.: 2.49      Length: 218
Score: 60.50      Matches: 17
Percent Similarity: 56.76%      Conservative: 4
Best Local Similarity: 45.95%      Mismatches: 11
Query Match: 29.09%      Indels: 5
DB: 3      Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-167-364-25 (1-218)

OY      3 GGGTGATCTTTTGCCCTTCAGAGATCTTTTCATCTTT-----GCAGGACTTCT 53
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      121 GlyTrpIleSerLeuTrpIleGlyPheSerPheIleMetPheThrSerAlaIleGlySerIu 140

OY      54 GGGGCCGGA---GTATGTAAGACTCTGGGCTCTGTGTGTGCTGAGTGG 101
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      141 GlyAlaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGluGluPhe 156

RESULT 5
US-09-439-897-4
; Sequence 4, Application US/09439897
; Patent No. 627558
; GENERAL INFORMATION:
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 95-1263-C
; CURRENT APPLICATION NUMBER: US/09/439,897
; NUMBER OF SEQ ID NOS: 65
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-897-4

Alignment Scores:
Pred. No.: 2.49      Length: 218
Score: 60.50      Matches: 17
Percent Similarity: 56.76%      Conservative: 4

```

Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
DB: 4 Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-439-897-4 (1-218)

QY 3 GGGTGGATCTTTGCGTGCAGGATCTTTTCATCTT-----GCAGGGACTTCT 53
DB 121 GYTRPILeserLeuTrpIysGlyPheSerPheIleMetPheThrSeraIaGlySerGlu 140
QY 54 GGGGCCGGA---GTATGTAAACTCTGCTGCTGTGTGCTGCTGAGTG 101
DB 141 GYlaIaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGluGluPhe 156

RESULT 6
US-09-589-927-6
; Sequence 6, Application US/09589927
; Patent No. 6432706
; GENERAL INFORMATION:

APPLICANT: University of Kansas Medical Center
TITLE OF INVENTION: The use of isolated domains of Type IV Collagen to
FILE REFERENCE: 945251
CURRENT APPLICATION NUMBER: US/09/589,927
CURRENT FILING DATE: 2000-06-07
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 268
TYPE: PRT
ORGANISM: Human
US-09-589-927-6

Alignment Scores:
Pred. No.: 2.57 Length: 268
Score: 60.50 Matches: 17
Percent Similarity: 56.76% Conservative: 4
Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
DB: 4 Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-589-927-6 (1-268)

QY 3 GGGTGGATCTTTGCGTGCAGGATCTTTTCATCTT-----GCAGGGACTTCT 53
DB 171 GYTRPILeserLeuTrpIysGlyPheSerPheIleMetPheThrSeraIaGlySerGlu 190
QY 54 GGGGCCGGA---GTATGTAAACTCTGCTGCTGTGTGCTGCTGAGTG 101
DB 191 GYlaIaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGluGluPhe 206

RESULT 7
US-09-277-665-6
; Sequence 6, Application US/09277665
; Patent No. 6440729
; GENERAL INFORMATION:
APPLICANT: University of Kansas Medical Center
TITLE OF INVENTION: The use of isolated domains of Type IV Collagen to
FILE REFERENCE: 94525-1
CURRENT APPLICATION NUMBER: US/09/277,665
CURRENT FILING DATE: 1999-03-26
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 6
LENGTH: 268
TYPE: PRT
ORGANISM: Human
US-09-277-665-6

Alignment Scores:
Pred. No.: 2.57 Length: 268
Score: 60.50 Matches: 17

Percent Similarity: 56.76% Conservative: 4
Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
DB: 4 Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-277-665-6 (1-268)

QY 3 GGGTGGATCTTTGCGTGCAGGATCTTTTCATCTT-----GCAGGGACTTCT 53
DB 171 GYTRPILeserLeuTrpIysGlyPheSerPheIleMetPheThrSeraIaGlySerGlu 190
QY 54 GGGGCCGGA---GTATGTAAACTCTGCTGCTGTGTGCTGCTGAGTG 101
DB 191 GYlaIaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGluGluPhe 206

RESULT 8
US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:

APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399,889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 24
LENGTH: 471
TYPE: PRT
ORGANISM: Calf
US-08-399-889-24

Alignment Scores:
Pred. No.: 2.81 Length: 471
Score: 60.50 Matches: 17
Percent Similarity: 56.76% Conservative: 4
Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
DB: 2 Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x US-08-399-889-24 (1-471)

QY 3 GGGTGGATCTTTGCGTGCAGGATCTTTTCATCTT-----GCAGGGACTTCT 53
DB 374 GYTRPILeserLeuTrpIysGlyPheSerPheIleMetPheThrSeraIaGlySerGlu 393
QY 54 GGGGCCGGA---GTATGTAAACTCTGCTGCTGTGTGCTGCTGAGTG 101
DB 394 GYlaIaGlyGlnAlaLeuAlaSerProGlySer---CysLeuGluGluPhe 409

RESULT 9
US-09-167-364-24
; Sequence 24, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263B
CURRENT APPLICATION NUMBER: US/09/167,364
CURRENT FILING DATE: 1998-10-07
EARLIER APPLICATION NUMBER: 08/399889
EARLIER FILING DATE: 1995-03-07
NUMBER OF SEQ ID NOS: 25
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 24
LENGTH: 471

[illegible]

| | |
|-----------------------------|------------|
| EARLIER FILING DATE: | 1997-06-06 |
| EARLIER APPLICATION NUMBER: | 60/056,886 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,877 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,889 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,893 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,630 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,878 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,662 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,872 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,903 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,888 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,879 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,880 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,894 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,911 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,636 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,874 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,910 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,864 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,631 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,845 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/056,892 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/057,761 |
| EARLIER FILING DATE: | 1997-08-22 |
| EARLIER APPLICATION NUMBER: | 60/047,593 |
| EARLIER FILING DATE: | 1997-05-23 |
| EARLIER APPLICATION NUMBER: | 60/047,586 |
| EARLIER FILING DATE: | 1997-05-23 |
| EARLIER APPLICATION NUMBER: | 60/047,590 |
| EARLIER FILING DATE: | 1997-05-23 |
| EARLIER APPLICATION NUMBER: | 60/047,594 |
| EARLIER FILING DATE: | 1997-05-23 |
| EARLIER APPLICATION NUMBER: | 60/047,586 |
| EARLIER FILING DATE: | 1997-05-23 |
| EARLIER APPLICATION NUMBER: | 60/047,593 |
| EARLIER FILING DATE: | 1997-05-23 |
| EARLIER APPLICATION NUMBER: | 60/047,595 |
| EARLIER FILING DATE: | 1997-05-23 |
| EARLIER APPLICATION NUMBER: | 60/047,614 |
| EARLIER FILING DATE: | 1997-05-23 |
| EARLIER APPLICATION NUMBER: | 60/043,578 |
| EARLIER FILING DATE: | 1997-04-11 |
| EARLIER APPLICATION NUMBER: | 60/043,576 |
| EARLIER FILING DATE: | 1997-04-11 |

PRIOR APPLICATION DATA:
APPLICATION NUMBER: No. 5501969e
FILING DATE: No. 5501969e
ATTORNEY/AGENT INFORMATION:
NAME: FERRARO, GREGORY D.
REGISTRATION NUMBER: 36,134
REFERENCE/DOCKET NUMBER: 325800-95
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-994-1700
TELEFAX: 201-994-1744
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 AMINO ACIDS
TYPE: AMINO ACID
STRANDEDNESS:
TOPOLOGY: LINEAR
MOLECULE TYPE: PROTEIN
US-08-208-007A-13

Alignment Scores:
Pred. No.: 11.2 Length: 396
Score: 56.00 Matches: 12
Percent Similarity: 62.07% Conservative: 6
Best Local Similarity: 41.38% Mismatches: 11
Query Match: 26.92% Indels: 0
Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-08-208-007A-13 (1-396)

QY 4 GGTGATCTTTTGCCTTCAGGATCTTTTCATCTTTGAGGACTTCGGGCGGAG 63
||| |||||:||||| ||||| |||||:||||| :|||
Db 132 GlyGlnSerPheSerIleGlnTyrGlyThrGlySerLeuSerGlyIleIleGlyAlaAsp 151

QY 64 TATGTAAACCTCCTGGCTCTGTGTG 90
||| :||| ||||| |||
Db 152 GlnValSerValGlnGlyLeuThrVal 160

RESULT 15
US-09-032-523-9
Sequence 9, Application US/09032523
Patent No. 6232454
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl
APPLICANT: Baugh, Mariah
TITLE OF INVENTION: HUMAN PROTEINASE MOLECULES
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,523
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0479 US
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: Genbank
CLONE: 181994
US-09-032-523-9

Alignment Scores:
Pred. No.: 11.2 Length: 396
Score: 56.00 Matches: 12
Percent Similarity: 62.07% Conservative: 6
Best Local Similarity: 41.38% Mismatches: 11
Query Match: 26.92% Indels: 0
Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x US-09-032-523-9 (1-396)

QY 4 GGTGATCTTTTGCCTTCAGGATCTTTTCATCTTTGAGGACTTCGGGCGGAG 63
||| |||||:||||| ||||| |||||:||||| :|||
Db 132 GlyGlnSerPheSerIleGlnTyrGlyThrGlySerLeuSerGlyIleIleGlyAlaAsp 151

QY 64 TATGTAAACCTCCTGGCTCTGTGTG 90
||| :||| ||||| |||
Db 152 GlnValSerValGlnGlyLeuThrVal 160

Search completed: April 30, 2003, 14:39:10
Job time : 14 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 30, 2003, 14:34:46 ; Search time 27.5 Seconds
(without alignments)
1663.362 Million cell updates/sec

Title: US-09-513-999c-3792_COPY_51_161
Perfect score: 208
Sequence: 1 atgggtgcatcttctgcctt.....gcttgagtgagctctact 111

Scoring table:
BIOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09513999/runatc_30042003_143401_25604/app_query.fasta_1.263
-DB=SPTRMBL_21 -OFMT=fastan -SUFFIX=n2p -rspt -MINMATCH=0.1 -LOOPC=0
-LIST=45 -UNITS=bits -START=1 -END=1 -MATRIX=blonsum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=ptc -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-USER=US09513999 @CGN_1.1.138 @runatc_30042003_143401_25604 -NCP=6 -ICPU=3
-NO_XLPRY -NO_MMAP -LARGESOURT -NRG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteria:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match Length | DB ID | Description |
|------------|-------|--------------|--------------|--------------------|
| c 1 | 66 | 32.7 | 83 15 Q80822 | Q80822 human t-cel |

| | | | | |
|------|------|------|----------------|--------------------|
| c 2 | 63.5 | 31.4 | 547 4 Q9Y5A5 | Q9Y5A5 homo sapien |
| c 3 | 63.5 | 31.4 | 634 4 Q9NSM4 | Q9NSM4 homo sapien |
| c 4 | 60.5 | 29.1 | 203 6 Q28682 | Q28682 oryctolagus |
| c 5 | 60.5 | 29.1 | 210 6 Q28273 | Q28273 canis famli |
| c 6 | 60.5 | 30.0 | 644 5 Q9N4W3 | Q9N4W3 caenorhabdi |
| c 7 | 60 | 28.8 | 314 5 Q20146 | Q20146 caenorhabdi |
| c 8 | 60 | 29.7 | 533 11 P97320 | P97320 mus musculu |
| c 9 | 60 | 28.8 | 1145 10 Q9LF0 | Q9LF0 arabisdopsi |
| c 10 | 60 | 29.7 | 5170 5 Q17490 | Q17490 caenorhabdi |
| c 11 | 60 | 29.7 | 6694 5 Q17343 | Q17343 caenorhabdi |
| c 12 | 59.5 | 29.5 | 1167 5 Q9Y067 | Q9Y067 thelleria a |
| c 13 | 59 | 28.4 | 382 16 Q8UEA8 | Q8UEA8 agrobacteri |
| c 14 | 59 | 29.2 | 537 11 Q9QY56 | Q9QY56 mus musculu |
| c 15 | 58.5 | 28.1 | 230 11 Q63122 | Q63122 rattus norv |
| c 16 | 58.5 | 28.1 | 231 15 Q56309 | Q56309 wallaye epl |
| c 17 | 58.5 | 29.0 | 283 4 Q8MM87 | Q8MM87 homo sapien |
| c 18 | 58.5 | 29.0 | 583 5 Q9YVH5 | Q9YVH5 drosophila |
| c 19 | 58 | 28.7 | 197 10 Q9SZ01 | Q9SZ01 arabisdopsi |
| c 20 | 58 | 27.9 | 257 12 Q9J2L8 | Q9J2L8 macaca mula |
| c 21 | 58 | 27.9 | 299 12 Q9MR74 | Q9MR74 macaca mula |
| c 22 | 58 | 27.9 | 499 17 Q8RPF6 | Q8RPF6 methanosarc |
| c 23 | 58 | 28.7 | 1099 5 Q9YMS5 | Q9YMS5 drosophila |
| c 24 | 57.5 | 27.6 | 212 6 Q28567 | Q28567 ovis aries |
| c 25 | 57.5 | 28.5 | 364 4 Q9BY89 | Q9BY89 homo sapien |
| c 26 | 57.5 | 28.5 | 1511 4 Q75412 | Q75412 homo sapien |
| c 27 | 57.5 | 28.5 | 1587 4 Q00508 | Q00508 homo sapien |
| c 28 | 57 | 27.4 | 240 8 Q20933 | Q20933 simulum vi |
| c 29 | 57 | 28.2 | 302 13 Q319C3 | Q319C3 anguilla an |
| c 30 | 57 | 28.2 | 880 10 Q9SV03 | Q9SV03 arabisdopsi |
| c 31 | 57 | 28.2 | 924 10 Q949M4 | Q949M4 arabisdopsi |
| c 32 | 57 | 28.2 | 3179 12 Q8V2A4 | Q8V2A4 human herpe |
| c 33 | 56.5 | 27.2 | 203 6 Q29032 | Q29032 sus scrofa |
| c 34 | 56.5 | 27.2 | 212 6 Q28512 | Q28512 macaca mula |
| c 35 | 56.5 | 27.2 | 245 4 Q9NYC4 | Q9NYC4 homo sapien |
| c 36 | 56.5 | 27.2 | 342 10 Q9M9M8 | Q9M9M8 arabisdopsi |
| c 37 | 56 | 26.9 | 204 8 Q8SG48 | Q8SG48 heteropsilo |
| c 38 | 56 | 26.9 | 315 16 P72593 | P72593 synechocyst |
| c 39 | 56 | 26.9 | 363 4 Q9NY58 | Q9NY58 homo sapien |
| c 40 | 56 | 27.7 | 468 11 Q9CYA6 | Q9CYA6 mus musculu |
| c 41 | 56 | 27.7 | 468 11 Q31WQ1 | Q31WQ1 mus musculu |
| c 42 | 56 | 27.7 | 511 4 Q96K62 | Q96K62 homo sapien |
| c 43 | 56 | 27.7 | 554 5 Q9W4C1 | Q9W4C1 drosophila |
| c 44 | 56 | 26.9 | 648 5 Q9NMD7 | Q9NMD7 drosophila |
| c 45 | 56 | 26.9 | 701 5 Q9VJU4 | Q9VJU4 drosophila |

ALIGNMENTS

RESULT 1
ID Q80822 PRELIMINARY: PRT: 83 AA.
AC Q80822;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE Protein 11 xv.
GN XV.
OS Human T-cell leukemia virus type II (HTLV-II).
OC Viruses; Retrovird viruses; Retroviridae; Deltaretrovirus.
OX NCBITaxID=11909;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95297146; PubMed=753968;
RA Clininale V., D'Agostino D.M., Zotti L., Franchini G., Felber B.K.,
RA Chieco-Bianchi L.;
RT "Expression and characterization of proteins produced by mRNAs spliced
RT into the x region of the human T-cell leukemia/lymphotropic virus type
RT II.";
RL Virology 209:445-456(1995).
DR EMBL: L41675; AAA98638.1;
SQ SEQUENCE 83 AA; 8427 MW; C4615BC04D9FD684 CRC64;

Alignment Scores:

[illegible]

```

FT  NON_TER  210  210
SQ  SEQUENCE  210 AA; 23025 MW; 31119E4CA823633D CRC64;

Alignment Scores:
Pred. No.: 9.55 Length: 210
Score: 60.50 Matches: 17
Percent Similarity: 56.76% Conservative: 4
Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
DB: 6 Gaps: 3

US-09-513-999C-3792_COPY_51_161 (1-111) x Q28273 (1-210)
QY 3 GGGTGCATCTTTTGGCTTCGACAGATCTTTTTCATCTTT-----GCAGGACTTCT 53
|||||
Db 125 G1TrrpleserleutrrpysylpheserPhillewerPherthseralagysrlu 144
|||||

OY 54 GGGGCGCGGA---GTATGTAAACTTCGCGGCTCTGTGTGTGGCTTCAGGG 101
|||||
Db 145 G1YalaclyGlnalaleuAlaserProlyser---CysleuIduInphe 160
|||||

RESULT 6
O9N4W3 PRELIMINARY; PRT; 644 AA.
ID O9N4W3;
AC O9N4W3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DI 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Y46H3C.4 protein.
GN Y46H3C.4
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Stoneking T., Wohlmann P., Lennox S.;
RT "The sequence of C. elegans cosmid Y46H3C.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006776; AAF60626.1; -.
DR HSSP; P06786; 1BGW.
DR InterPro; IPR002205; DNA_topoisolv.
DR Pfam; PF00521; DNA_topoisolv; 1.
DR ProDom; PD000742; DNA_topoisolv; 1.
DR SMART; SM00434; TOP4c; 1.
SQ SEQUENCE 644 AA; 73898 MW; C530B213895D5958 CRC64;

Alignment Scores:
Pred. No.: 9.98 Length: 644
Score: 60.50 Matches: 17
Percent Similarity: 53.66% Conservative: 5
Best Local Similarity: 41.46% Mismatches: 14
Query Match: 29.95% Indels: 5
DB: 5 Gaps: 1

US-09-513-999C-3792_COPY_51_161 (1-111) x Q9N4W3 (1-644)
QY 108 AGAGCAGCCACTCAGGCGCACACAGAGACCGACGGAGTTTATCACTTCGGCCCCAGAG 49
|||||
|||||
|||||

```

```

Db 281 ArgThrSerThrSerProGluValIleTherGlnIupHeThyAspSerArgGlnIulys 300
QY 48 TCCCGCAAGATGAAAAAGATCTCTG-----CAAGCGAAAAAGATCCACC 4
Db 301 TTTLeuGlnArgLysGlnIupHeTherGlnIulysValLeuGlnIulysSerLysArgLeuThr 320
QY 3 CAT 1
Db 321 Asn 321

RESULT 7
Q20146 PRELIMINARY; PRT; 314 AA.
ID Q20146
AC Q20146;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 36.0 kDa protein.
GN F38B6.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodermidae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA W11Cox L.;
RT "The sequence of C. elegans cosmid F38B6.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; 040060; AAA81145.1; -.
DR InterPro; IPR002899; WRI/EB.
DR SMART; SM00289; WRI; 1.
KW Hypothetical protein.
SQ SEQUENCE 314 AA; 36010 MW; 6C5042D23E8038E6 CRC64;

Alignment Scores:
Pred. No.: 11.4 Length: 314
Score: 60.00 Matches: 11
Percent Similarity: 62.96% Conservative: 6
Best Local Similarity: 40.74% Mismatches: 6
Query Match: 28.85% Indels: 4
DB: 5 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x Q20146 (1-314)

QY 11 CTTTGGCTTGACGAGATCTTTTCATCTT-----TGCAGGAGCTCTGGCGC 58
Db 278 TTTLeuLysGlnArgLysGlnIupHeTherGlnIulysValLeuGlnIulysSerLysArgLeuThr 297
QY 59 CGGAGTATGTAAACTGCTGG 79
Db 298 ArgLysLeuGlnIulysAsnLeuTyr 304

RESULT 8
P97320 PRELIMINARY; PRT; 533 AA.
ID P97320
AC P97320;
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)

```

```

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Testis-specific protein, DDC8.
GN DDC8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CD1;
RA Catalano R.D., Vlad M., Kennedy R.C.;
RT "Differential display to identify and isolate novel genes expressed
RT during spermatogenesis.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y09878; CAA71005.1; -.
DR MGD; MGI:1929713; Ddc8.
SQ SEQUENCE 533 AA; 62039 MW; C7F75F456B24F52 CRC64;

Alignment Scores:
Pred. No.: 11.6 Length: 533
Score: 60.00 Matches: 11
Percent Similarity: 52.94% Conservative: 7
Best Local Similarity: 32.35% Mismatches: 16
Query Match: 29.70% Indels: 0
DB: 11 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x P97320 (1-533)

QY 107 GAGCAGCACCTGACGACACAGAGACCCAGATTTCATCTGCGGCCGAGAGT 48
Db 116 GtGluArgGlyArgArgGlnIupHeTherGlnIulysProLysSerArgLysGlnIulys 135
QY 47 CCTGCAAGATGAAAAAGATCTCTGCAAGCGAAAAAGATCCA 6
Db 136 ProCysSerGlnArgSerSerAlaLysAlaArgPro 149

RESULT 9
Q9LF10 PRELIMINARY; PRT; 1145 AA.
ID Q9LF10
AC Q9LF10;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Cellulose synthase catalytic subunit-like protein.
GN F2K13.60.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Kotani H.,
RA Tabata S., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391141; CAC01704.1; -.
DR InterPro; IPR005150; Cellulose_synth.
DR Pfam; PF03552; Cellulose_synth; 1.
SQ SEQUENCE 1145 AA; 128359 MW; 21A37FD050DC26BC CRC64;

Alignment Scores:
Pred. No.: 12 Length: 1145
Score: 60.00 Matches: 13
Percent Similarity: 57.14% Conservative: 3
Best Local Similarity: 46.43% Mismatches: 11
Query Match: 28.85% Indels: 1
DB: 10 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x Q9LF10 (1-1145)

```

```

DE 01-JUN-1998 (TREMBLrel. 05, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE UNC-44 ankyrins.
GN UNC-44.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N2;
RX MEDLINE=95263663; PubMed=7744957;
RA Otsuka A.J., Franco R., Yang B., Shim K., Tang L.Z., Zhang Y.,
RA Boontrakulpoonatawee P., Jayaprakash A., Hedgecock E., Wheaton V.I.,
RA Sobery A.;
RT "An ankyrin-related gene (unc-44) is necessary for proper axonal
RT guidance in Caenorhabditis elegans."
RL J. Cell Biol. 129:1081-1092(1995).
RN [2]
RP REVISIONS, AND SEQUENCE OF 6126-6994 FROM N.A.
RC STRAIN=N2;
RA Otsuka A.J.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U39847; AAB41827.1; -.
DR EMBL: U21733; AAB38384.1; -.
DR HSSP: P42773; 11HR.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000488; Death.
DR InterPro: IPR001360; GH_1.
DR InterPro: IPR002383; GLA_blood.
DR InterPro: IPR000906; ZUS.
DR Pfam: PF00023; ank; 24.
DR Pfam: PF00531; death; 1.
DR Pfam: PF00791; ZUS; 1.
DR PRINTS: PRO1415; ANKYRIN.
DR PRINTS: PRO0001; GLABLOOD.
DR SMART: SM00248; ANK; 21.
DR SMART: SM00005; DEATH; 1.
DR SMART: SM00218; ZUS; 1.
DR PROSITE: PS50088; ANK_REPEAT; 22.
DR PROSITE: PS50297; ANK_REC_REGION; 1.
DR PROSITE: PS50017; DEATH_DOMAIN; 1.
DR PROSITE: PS00572; GLYCOSTL_HYDROL_F1_1; UNKNOWN_2.
KW ANK repeat; Repeat.
SQ SEQUENCE 6994 AA; 775364 MW; 90CB449925D9923D CRC64;

Alignment Scores:
Pred. No.: 12.9 Length: 6994
Score: 60.00 Matches: 11
Percent Similarity: 51.85 Conservative: 3
Best Local Similarity: 40.74 Mismatches: 13
Query Match: 29.70% Indels: 0
DB: 5 Gaps: 0

US-09-513-999C-3792_COPY_51_161 (1-111) x Q17343 (1-6994)
QY 107 GAGCAGCCACTCAGGCACACAGAGACCAGAGTTCATTAAGTCCGCGCCCAAGAGT 48
||| |||||:: |||::: ||| ||| ||| |||
Db 2796 GUGBGRFRLGGLGHTSGILYSGILUTHTGILGULPHEASRHSERHSRPROSLUBR 2815
QY 47 CCGTCGCAAGATGAAAAAGAA 27
||| ||||| |||
Db 2816 Provalleusergiluygsiu 2822

RESULT 12
Q9Y067 PRELIMINARY; PRT; 1167 AA.
AC Q9Y067;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE TASHAT2 protein.
EN TASHAT2.

```


US-09-513-999c-3792_COPY_51_161 (1-111) x 090XS6 (1-537)

QY 106 AGCAGCCACTTCAGCAGACAGACAGCCAGGAGTTTACATPACTCCGCCGCCAGAGTC 47

Db 484 SerSerHisSer-----SerLeuAlaHisLeuProAlaProVal 497

QY 46 CCTGCAAAAGATGAAAAAGATCCTGCAGGCAAGATCACCACA 2

Db 498 ProProArgIlnsSerSerProLeuProIlysLeuProPro 512

RESULT 15

Q63122 ID Q63122 PRELIMINARY; PRT; 230 AA.

AC Q63122;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE Alpha-3 type IV collagen (Fragment).

IN COL4A3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY CORTEX;

RX MEDLINE-98210005; PubMed-9550634;

RA Ryan J.J., Katbama I., Mason P.J., Pusey C.D., Turner A.N.;

RT "Sequence analysis of the 'Goodpasture antigen' of mammals.,"

RL Nephrol. Dial. Transplant. 13:602-607(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-SPRAGUE-DAWLEY; TISSUE-KIDNEY CORTEX;

RA Turner N.;

RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL: LA7281; AAB72238.2;

DR InterPro: IPR001442; ProcollagenC4.

DR Pfam: PF01413; C4; 2.

DR ProDom: PD003923; ProcollagenC4; 1.

DR SMART: SM00111; C4; 2.

DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.

KW Collagen.

FT NON_TER 1

FT NON_TER 230

SQ SEQUENCE 230 AA; 25398 MW; 29549E5314CC056 CRC64;

Alignment Scores:

Ref. No.:

Score: 18.2

Percent Similarity: 58.50

Best Local Similarity: 56.76%

Query Match: 40.54%

Db: 28.12%

Length: 230

Matches: 15

Conservative: 6

Mismatches: 11

Indels: 5

Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x Q63122 (1-230)

QY 3 GGGTGGATCTTTGGCTTCAGAGATCTTTTCATCTT-----GCAGCGACTTCT 53

Db 133 GlyTPrValSerLeuTrpIysGlyPheSerPheValMetPheTrhSerAlaGlySerIlu 152

QY 54 GGGGCCGCGA--GTAATGTAACCTCTGGCTCTGTGTGTGTGCTGAGTGG 101

Db 153 GlyAlaGlyGlnAlaLeuAlaSerProGlySer---CysIeuGluGluPhe 168

Search completed: April 30, 2003, 14:36:15
Job time : 32.5 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 30, 2003, 14:34:42 ; Search time 9 Seconds
(Without alignments)
1023.083 Million cell updates/sec

Title: US-09-513-999c-3792_COPY_51_161
Perfect score: 208
Sequence: 1 atggatgacatcttgcctt.....gcctgagtgagctctact 111

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09513999/umat_30042003_143401_25593/app_query.fasta.1.263
-DB=swisprot.40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOEXT=0
-LOOEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09513999.ecgn.1.1.26.umat.30042003.143401.25593 -NCPU=6 -ICPU=3
-NO_XLPY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN.TIMEOUT=30 -THRAADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FEAPOP=6 -FEAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwisProt.40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| c 1 | 63.5 | 31.4 | 648 | 1 | 2202_HUMAN |
| c 2 | 60.5 | 29.1 | 471 | 1 | CA34_BOVIN |
| c 3 | 59 | 29.2 | 1297 | 1 | SOS2_MOUSE |
| c 4 | 58.5 | 29.0 | 668 | 1 | SCCL_HUMAN |
| c 5 | 58 | 27.9 | 391 | 1 | CATE_CAVPO |
| c 6 | 57 | 28.2 | 3149 | 1 | TEGU_EBV |
| c 7 | 56.5 | 28.0 | 938 | 1 | CAFA_HUMAN |
| c 8 | 56.5 | 27.2 | 1670 | 1 | CA34_HUMAN |
| c 9 | 56 | 26.9 | 396 | 1 | CATE_HUMAN |
| c 10 | 55.5 | 26.7 | 147 | 1 | THY_HUMAN |
| c 11 | 55.5 | 26.7 | 687 | 1 | VS41_GIALA |
| c 12 | 54.5 | 27.0 | 113 | 1 | N12B_MEDSA |
| c 13 | 54.5 | 26.2 | 150 | 1 | THY_PIG |
| c 14 | 54.5 | 26.2 | 397 | 1 | UTR_RAT |
| c 15 | 54.5 | 26.2 | 481 | 1 | PGLI_JUNIN |
| c 16 | 54.5 | 27.0 | 1520 | 1 | TOP2_CAHEL |
| c 17 | 54 | 26.0 | 387 | 1 | CYB_PODAN |
| c 18 | 54 | 26.7 | 731 | 1 | MR11_CORCI |

| | | | | | | |
|------|------|------|------|---|------------|--------------------|
| c 19 | 53.5 | 26.5 | 634 | 1 | HNFA_CHICK | 090867 gallus gall |
| c 20 | 53.5 | 26.5 | 986 | 1 | AC15_DROME | P33600 drosophila |
| c 21 | 53 | 26.2 | 118 | 1 | VEA_HPV13 | 002665 human papil |
| c 22 | 53 | 26.2 | 346 | 1 | GCP_BORBU | 051710 borrelia bu |
| c 23 | 53 | 25.3 | 397 | 1 | CATE_MOUSE | P70269 mus musculu |
| c 24 | 53 | 25.5 | 398 | 1 | CATE_RAT | P16228 rattus norv |
| c 25 | 53 | 26.2 | 716 | 1 | PEP_DROME | P41073 drosophila |
| c 26 | 52.5 | 25.2 | 147 | 1 | THY_BOVIN | 046375 bos taurus |
| c 27 | 52.5 | 25.2 | 389 | 1 | UTL_HUMAN | Q13336 homo sapien |
| c 28 | 52.5 | 25.2 | 531 | 1 | PNX1_SCHPO | 009752 schizosach |
| c 29 | 52.5 | 26.0 | 768 | 1 | CNIC_RAT | Q63421 rattus norv |
| c 30 | 52 | 25.0 | 204 | 1 | GTR5_BOVIN | P58353 bos taurus |
| c 31 | 52 | 25.0 | 587 | 1 | LACP_KIUUA | P07921 kluyveromyc |
| c 32 | 52 | 25.7 | 1581 | 1 | LMG3_MOUSE | Q92538 mus musculu |
| c 33 | 52 | 25.7 | 1859 | 1 | GFL1_HUMAN | Q92538 mus sapien |
| c 34 | 51.5 | 24.8 | 147 | 1 | THY_RAT | P02767 rattus norv |
| c 35 | 51.5 | 25.5 | 248 | 1 | DJ_DROME | 001352 drosophila |
| c 36 | 51 | 25.2 | 50 | 1 | RS30_AERPE | Q9969 aeropyrum p |
| c 37 | 51 | 25.2 | 269 | 1 | HXB5_HUMAN | P03067 homo sapien |
| c 38 | 51 | 25.2 | 269 | 1 | HXB5_MOUSE | P03079 mus musculu |
| c 39 | 51 | 25.2 | 318 | 1 | KHSE_VIBCH | Q9KPK4 vibrio chol |
| c 40 | 51 | 24.5 | 354 | 1 | NOV_MOUSE | Q64299 mus musculu |
| c 41 | 51 | 24.5 | 452 | 1 | AD11_XENLA | Q98823 xenopus lae |
| c 42 | 51 | 25.2 | 492 | 1 | CAT3_SOYBN | Q48560 glycine max |
| c 43 | 51 | 24.5 | 679 | 1 | SYGB_BACSU | P54381 bacillus su |
| c 44 | 51 | 24.5 | 866 | 1 | NSGA_KLEPN | 006457 klebsiella |
| c 45 | 51 | 25.2 | 1119 | 1 | RPOB_THEAO | Q9KXU7 thermus aqu |

ALIGNMENTS

RESULT 1
ID 2202_HUMAN STANDARD; PRT; 648 AA.
AC 095125; Q9H1B9.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 202.
GN ZNF202.
GN Homo sapiens (human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND VARIANT VAL-154.
RC TISSUE=Testis;
RX MEDLINE=99009320; PubMed=9790754;
RA Monaco C., Helmer Citterich M., Caprioli E., Vorechovsky I., Russo G.,
RT "Molecular cloning and characterization of ZNF202: A new gene at
RT 11q23.3 encoding testis-specific zinc finger proteins.";
RL Genomics 52:358-362(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Langmann T., Poroch-Ozcurmez M., Helmerl S., Andrikovics H.,
RA Schmitz G.;
RT "Genomic sequence analysis of the ZNF202 gene: relevance for lipid
RT parameters.";
RN Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=20283591; PubMed=10748193;
RA Wagner S., Hess M.A., Ormonde-Hanson P., Malandro J., Hu H., Chen M.,
RA Kehrle R., Frodsham D., Schumacher C., Beluch M., Honer C.,
RT "A broad role for the zinc finger protein ZNF202 in human lipid
RT metabolism.";
RL J. Biol. Chem. 275:15685-15690(2000).
CC -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR THAT BINDS TO ELEMENTS FOUND
CC PREDOMINANTLY IN GENES THAT PARTICIPATE IN LIPID METABOLISM. AMONG
CC ITS TARGETS ARE STRUCTURAL COMPONENTS OF LIPOPROTEIN PARTICLES
CC (APOLIPOPROTEINS AIV, CIII, AND E), ENZYMES INVOLVED IN LIPID

```
CC CC PROCESSING (LIPOPROTEIN LIPASE, LECTHIN CHOLESTERYL ESTER  
CC TRANSFERASE). AND SEVERAL GENES INVOLVED IN PROCESSES RELATED TO  
CC ENERGY METABOLISM AND VASCULAR DISEASE.  
CC -1- SUBUNIT: INTERACTS WITH SOD1.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA/1 AND BETA/2 (SHOWN HERE);  
CC ARE PRODUCED BY ALTERNATE SPLICING.  
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. ALSO EXPRESSED IN  
CC BREAST CARCINOMA CELL LINES.  
CC -1- SIMILARITY: CONTAINS 1 SCAN BOX.  
CC -1- SIMILARITY: CONTAINS 1 KRAB DOMAIN.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC The European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see http://www.isb.ch/announce/  
CC or send an email to license@isb.sib.ch).  
-----  
DR EMBL AF027219; AAC79941.1; -;  
DR EMBL AF027218; AAC79940.1; -;  
DR EMBL AJ276177; CAC21447.1; -;  
DR EMBL AJ276178; CAC21447.1; JOINED.  
DR EMBL AJ276179; CAC21447.1; JOINED.  
DR EMBL AJ276180; CAC21447.1; JOINED.  
DR EMBL AJ276181; CAC21447.1; JOINED.  
DR EMBL AJ276182; CAC21447.1; JOINED.  
DR HSSP: P08047; ISP2.  
DR Genew: HGNC:12994; ZNF202.  
DR MIM: 603450; -;  
DR InterPro: IPR001909; KRAB.  
DR InterPro: IPR003309; Treg_SCAN.  
DR InterPro: IPR000822; ZnF_C2H2.  
DR Pfam: PF000096; zf-C2H2; 8.  
DR Pfam: PF013352; KRAB; 1.  
DR Pfam: PF02023; SCAN; 1.  
DR PRINTS: PR00048; ZINC_FINGER.  
DR PRODOM: PD000003; ZnF_C2H2; 4.  
DR SMART: SMO0349; KRAB; 1.  
DR SMART: SMO0431; LER; 1.  
DR SMART: SMO0355; znf.C2H2; 8.  
DR PROSITE: PS50805; KRAB; 1.  
DR PROSITE: PS50804; SCAN_BOX_1.  
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 8.  
DR PROSITE: PSS0157; ZINC_FINGER_C2H2_2; 8.  
KW Transcription regulation; zinc-finger; Metal-binding; DNA-binding;  
KW Nuclear protein; Repressor; Repeat; Alternative splicing;  
KW Polymorphism.  
FT DOMAIN 46 127 SCAN_BOX.  
FT FT 237 308 KRAB.  
FT DOMAIN 397 447 ZINC_FINGERS.  
FT FT 447 447 ZINC_FINGERS.  
FT DOMAIN 481 643 ZINC_FINGERS.  
FT ZN_FING 397 419 C2H2-TYPE.  
FT ZN_FING 425 447 C2H2-TYPE.  
FT ZN_FING 481 503 C2H2-TYPE.  
FT ZN_FING 509 531 C2H2-TYPE.  
FT ZN_FING 537 559 C2H2-TYPE.  
FT ZN_FING 565 587 C2H2-TYPE.  
FT ZN_FING 591 615 C2H2-TYPE.  
FT ZN_FING 621 643 C2H2-TYPE.  
FT VARSPLOC 1 224 MISSING (IN ISOFORM ALPHA).  
FT VARIANT 154 154 A->V.  
FT /FMTD-VAR_007818.  
FT CONFLICT 205 205 MISSING (IN REF. 2).  
FT CONFICT 278 278 S->AA (IN REF. 2).  
SQ SEQUENCE 648 AA; 74692 MW; 357174FE3BFBFAE CRC64;
```

```

Query Match:      31.44% Indels:      5
DB:               1 Gaps:           3

US-09-513-999C-3792_COPY_51_161 (1-111) x Z202_HUMAN (1-648)

OY    110 GTACAGCAGCCATCTGCACACACAGACGACG-----AGTTTTCATACTCGGC 57
Db     300 IlegInguPro---GIingIuthrGIingluPIroGLIUleusSerherThyrthrGly 318
OY      56 CCCAAGATCCCTGCAAGAATGAAMAAAGATTCTGCAGGCCAAAAGAT 9
Db     319 AspArgSer-----LysAspGIugluGlucylcSLeugluGlnGlnusp 332

RESULT 2
CA34_BOVIN STANDARD; PRT; 471 AA.
ID CA34_BOVIN Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Collagen alpha 3(IV) chain (Fragment).
GN COL4A3.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidia; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=91093146; PubMed=1985905;
RA Morrison K.E., Germino G.G., Reeders S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
RL J. Biol. Chem. 266:34-39(1991)."
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RA Gunwar S., Sauz J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RL J. Biol. Chem. 265:5466-5469(1990)."
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=88330844; PubMed=3417661;
RA Sauz J., Wieslander J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
RL J. Biol. Chem. 263:13374-13380(1988)."
RN [4]
RP SEQUENCE OF 227-244.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowsky R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
membrane collagen."
RT J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
MESHWORK TOGETHER WITH LAMININS, PROTEOLYCAN AND ENACTIN/N
NIDOSEN.
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH
CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
TRIPLE-HELICAL 7S DOMAIN.
CC -1- PTM: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

```

```

CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M63139; AAA62708.1; -.
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR001442; ProcollagenC4.
CC Pfam: PF01391; Collagen; 4.
CC ProDom: PD003923; ProcollagenC4; 2.
CC SMART: SM00111; C4; 2.
CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON_TER 1 238 TRIPLE-HELICAL REGION.
FT DOMAIN 239 471 NONHELICAL REGION (NC1).
FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 232 232 HYDROXYLATION.
FT MOD_RES 238 238 HYDROXYLATION.
FT DISULFID 261 352 OR 349 (BY SIMILARITY).
FT DISULFID 294 349 OR 352 (BY SIMILARITY).
FT DISULFID 306 312 BY SIMILARITY.
FT DISULFID 371 466 OR 463 (BY SIMILARITY).
FT DISULFID 405 463 OR 466 (BY SIMILARITY).
FT DISULFID 417 423 BY SIMILARITY.
FT CONFLICT 253 253 S -> Y (IN REF. 3).
SQ SEQUENCE 471 AA; 47585 MW; C03B66F1AE7008DE CRC64;

Alignment Scores:
Pred. No.: 5.2 Length: 471
Score: 60.50 Matches: 17
Percent Similarity: 56.76% Conservative: 4
Best Local Similarity: 45.95% Mismatches: 11
Query Match: 29.09% Indels: 5
DB: 1 Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x CA34_BOVIN (1-471)
QY 3 GGGTGGATCTTTTGCCTTGACGAGATCTTTTCATCTTT-----GCAGGAGACTTCT 53
DB 374 GYTPRleSerleuTrpIysGlyPheSerPheIleMetPheThrSerIaGlySerGlu 393
QY 54 GGGGCCCGGA---GTATGTAAACTCCTGGCTCTGTGTGTCGCTGAGTGG 101
DB 394 GYIaGlyGlnAlaIleuAlaSerProGlySer---CysIeuGluGluPhe 409

RESULT 3
SOS2_MOUSE
ID SOS2_MOUSE STANDARD: PRT; 1297 AA.
AC 002384;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Son of sevenless protein homolog 2 (SOS-2) (mSOS-2) (Fragment).
GN SOS2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss; TISSUE=Eye;
RX MEDLINE=92335328; PubMed=1631150;

```

```

RA Bowtell D.D., Fu P., Simon M.A., Senior P.V.;
RT "Identification of murine homologues of the Drosophila son of
RT sevenless gene: potential activators of ras.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6511-6515(1992).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP (BY
CC SIMILARITY).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MOST EMBRYONIC AND ADULT TISSUES.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOG (DH) DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RAS-GEF DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z11664; CAA77732.1; -.
CC PIR: S25714; S25714.
CC HSSP: O62245; 1PMS.
CC MGD: MGI:98355; Sos2.
CC InterPro: IPR001331; GDS_CDC24.
CC InterPro: IPR001849; PH.
CC InterPro: IPR000651; RasGEFN.
CC InterPro: IPR001895; RasGEF_CDC25.
CC InterPro: IPR000213; RhogEF.
CC Pfam: PF00169; PH; 1.
CC Pfam: PF00617; RasGEF; 1.
CC Pfam: PF00618; RasGEFN; 1.
CC Pfam: PF00621; RhogEF; 1.
CC SMART: SM00233; PH; 1.
CC SMART: SM00147; RasGEF; 1.
CC SMART: SM00229; RasGEFN; 1.
CC SMART: SM00325; RhogEF; 1.
CC PROSITE: PS00741; DH_1; FALSE_NEG.
CC PROSITE: PS00720; DH_2; 1.
CC PROSITE: PS00720; GDS_CDC25; 1.
CC PROSITE: PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor.
FT NON_TER 1 1
FT DOMAIN 164 353 DH.
FT DOMAIN 407 510 PH.
FT DOMAIN 740 926 RAS-GEF.
FT DOMAIN 1145 1148 POLY-PRO.
FT DOMAIN 1164 1173 POLY-PRO.
SQ SEQUENCE 1297 AA; 14846 MW; 3A53F8515BF9D625 CRC64;

Alignment Scores:
Pred. No.: 8.17 Length: 1297
Score: 59.00 Matches: 13
Percent Similarity: 48.57% Conservative: 4
Best Local Similarity: 37.14% Mismatches: 12
Query Match: 29.21% Indels: 6
DB: 1 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x SOS2_MOUSE (1-1297)
QY 106 AGCAGCCTCAGGACACACAGAGACCCAGAGTTTACATCTCCGGCCCAAGAGTC 47
DB 1244 SerSerHisSer-----SerIeuAlaHisIeuProAlaProProVal 1257
QY 46 CCTGCAAGATGAAAAAGATCTCTGCAGGCAAAAGATCCACCA 2
DB 1258 ProProArGlnAAsnSerSerProIeuLeuProIysIeuProPro 1272

RESULT 4
SCEN_HUMAN
ID SCEN_HUMAN STANDARD: PRT; 668 AA.
AC O95171.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Scellin.
 GN SCEL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 443-457 AND 635-650.
 RC TISSUE=foreskin;
 RX MEDLINE=99030435; PubMed=9813070;
 RA Chompland M.-F., Burgess R.E., Jin W., Baden H.P., Olson P.F.;
 RT "CDNA cloning and characterization of scellin, a LIM domain protein
 of the keratinocyte cornified envelope.";
 RL J. Biol. Chem. 273:31547-31554(1998).
 CC -1- FUNCTION: May function in the assembly or regulation of proteins
 in the cornified envelope. The LIM domain may be involved in
 homotypic or heterotypic associations and may function to localize
 scellin to the cornified envelope.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. May become cross-linked to
 membrane proteins by transglutaminase.
 CC -1- TISSUE SPECIFICITY: Highly expressed in esophagus. It is also
 expressed in keratinocytes, amniotic tissue, foreskin stratum
 spinosum and stratum granulosum, hair follicle and nail.
 CC -1- SIMILARITY: CONTAINS 1 LIM DOMAIN. THE LIM DOMAIN BINDS 2 ZINC
 IONS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; AF045941; AAC78461.1; -;
 DR Genew; HGNC:10573; SCEL.
 DR MIM; 604112; -;
 DR InterPro; IPR001781; LIM.
 DR ProDom; PD000094; LIM; 1.
 DR SMART; SM00132; LIM; 1.
 DR PROSITE; PS00478; LIM_DOMAIN_1; FALSE_NEG.
 DR PROSITE; PSS0023; LIM_DOMAIN_2; 1.
 KW LIM domain; Metal-binding; Zinc; Repeat.
 FT DOMAIN 599 665 16 X APPROXIMATE TANDEM REPEATS.
 FT 231 243 1.
 FT REPEAT 231 246 1.
 FT REPEAT 247 266 2.
 FT REPEAT 267 286 3.
 FT REPEAT 287 306 4.
 FT REPEAT 307 326 5.
 FT REPEAT 327 346 6.
 FT REPEAT 347 366 7.
 FT REPEAT 367 386 8.
 FT REPEAT 387 406 9.
 FT REPEAT 407 426 10.
 FT REPEAT 427 445 11.
 FT REPEAT 446 464 12.
 FT REPEAT 465 484 13.
 FT REPEAT 485 503 14.
 FT REPEAT 504 523 15.
 FT REPEAT 524 543 16.
 SO SEQUENCE 668 AA; 75297 MW; 612BC3686DBA9FB0 CRC64;

 Alignment Scores:
 Pred. No.: 9 48
 Score: 58.50
 Percent Similarity: 56.108
 Best Local Similarity: 39.028
 Query Match: 28.968
 DB: 1
 Gaps: 3
 Length: 668
 Matches: 16
 Conservative: 7
 Mismatches: 11
 Indels: 7
 US-09-513-999c-3792_COPY_51_161 (1-111) x SCEL_HUMAN (1-668)

OY 106 AGCAGCCACTCAGGACACA---CACAGAGACCCAGAGACTTTACATCTCCGCCCCAGAA 50
 DB 174 SerSerSerThrGlyThrArgArgArgGlyProGlyVal---HisProProIleProPro 192
 OY 49 GTCCCGCAAGAATGAAAAAGATCCTGCAGAG-----CAAAAGATCCAC 5
 DB 193 LysProSerSerProValSerSerProValSerProValSerProValSerProValSer 212
 OY 4 CCA 2
 DB 213 Pro 213

 RESULT 5
 CATE_CAVPO
 ID CATE_CAVPO STANDARD; PRT; 391 AA.
 AC P25796;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cathepsin E precursor (EC 3.4.23.34).
 GN CTSE.
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystriocognathi; Cavidae; Cavia.
 OX NCBI_TaxID=10141;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=92355614; PubMed=1644829;
 RA Kageyama T., Ichinose M., Tsukada S., Miki K., Kurokawa K., Koiwai O.,
 RA Tanji M., Yakabe E., Althaus S.B., Takahashi K.;
 RT "Gastric procathepsin E and progastricsin from guinea pig.
 RT Purification, molecular cloning of cDNAs, and characterization of
 RT enzymatic properties, with special reference to procathepsin E.";
 RL J. Biol. Chem. 267:16450-16459(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX TISSUE=gastric mucosa;
 RX MEDLINE=96073637; PubMed=8540321;
 RA Kageyama T., Ichinose M., Miki K., Moriyama A., Yonezawa S.,
 RA Tanji M., Althaus S.B., Takahashi K.;
 RT "Isolation, characterization, and structure of procathepsin E and
 RT cathepsin E from the gastric mucosa of guinea pig.";
 Adv. Exp. Med. Biol. 362:211-221(1995).
 CC -1- FUNCTION: DUE TO ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
 CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
 specificity.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL; M88653; AAA37052.1; -;
 DR EMBL; S80547; AAB35844.1; -;
 DR PIR; A43356; A43356.
 DR HSSP; P00794; 4CMS.
 DR MEROPS; A01.010; -;
 DR InterPro; IPR001461; Aspartaseal.
 DR InterPro; IPR001969; Aspartase_site.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 18
 FT PROPEP 19 53
 FT CHAIN 54 391
 CATEPSIN E.

RA Bernal D., Quiñones S., Saus J.;
 RT "The human mRNA encoding the Goodpasture antigen is alternatively
 RT spliced.";
 RL J. Biol. Chem. 268:12090-12094(1993).
 RN (11)
 RP VARIANT PRO-1474,
 RX MEDLINE=95078827; PubMed=7987301;
 RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.M.J., Schroeder C.H.,
 RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
 RA Reekers S.T., Smeets H.J.M.;
 RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
 RT recessive Alport syndrome.";
 RL Hum. Mol. Genet. 3:1269-1273(1994).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/
 CC NIDOGIN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2/V AND
 CC 3/L5; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
 CC C-TERMINAL NC1 DOMAINS.
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
 CC COLLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
 CC COCHLEA, LUNG AND BRAIN.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLET REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: THE ALTERNATIVE SPLICED FORK V CONTAINS AN ADDITIONAL
 CC N-LINKED GLYCOSYLATION SITE.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- PTM: Phosphorylated by the Goodpasture antigen-binding protein.
 CC -1- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE
 CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
 CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
 CC -1- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
 CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
 CC MALES AND FEMALES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X80031; CAA56335.1; -;
 DR EMBL: AJ288487; CAC36101.1; -;
 DR EMBL: AJ288488; CAC36101.1; JOINED.
 DR EMBL: AJ288489; CAC36101.1; JOINED.
 DR EMBL: AJ288490; CAC36101.1; JOINED.
 DR EMBL: AJ288491; CAC36101.1; JOINED.
 DR EMBL: AJ288492; CAC36101.1; JOINED.
 DR EMBL: AJ288493; CAC36101.1; JOINED.
 DR EMBL: AJ288494; CAC36101.1; JOINED.
 DR EMBL: AJ288495; CAC36101.1; JOINED.
 DR EMBL: AJ288496; CAC36101.1; JOINED.
 DR EMBL: AJ288497; CAC36101.1; JOINED.
 DR EMBL: AJ288498; CAC36101.1; JOINED.
 DR EMBL: AJ288499; CAC36101.1; JOINED.
 DR EMBL: AJ288500; CAC36101.1; JOINED.

DR EMBL: AJ288501; CAC36101.1; JOINED.
 DR EMBL: AJ288502; CAC36101.1; JOINED.
 DR EMBL: AJ288503; CAC36101.1; JOINED.
 DR EMBL: AJ288504; CAC36101.1; JOINED.
 DR EMBL: AJ288505; CAC36101.1; JOINED.
 DR EMBL: AJ288506; CAC36101.1; JOINED.
 DR EMBL: AJ288507; CAC36101.1; JOINED.
 DR EMBL: AJ288508; CAC36101.1; JOINED.
 DR EMBL: AJ288509; CAC36101.1; JOINED.
 DR EMBL: AJ288510; CAC36101.1; JOINED.
 DR EMBL: AJ288511; CAC36101.1; JOINED.
 DR EMBL: AJ288512; CAC36101.1; JOINED.
 DR EMBL: AJ288513; CAC36101.1; JOINED.
 DR EMBL: AJ288514; CAC36101.1; JOINED.
 DR EMBL: AJ288515; CAC36101.1; JOINED.
 DR EMBL: AJ288516; CAC36101.1; JOINED.
 DR EMBL: AJ288517; CAC36101.1; JOINED.
 DR EMBL: AJ288518; CAC36101.1; JOINED.
 DR EMBL: AJ288519; CAC36101.1; JOINED.
 DR EMBL: AJ288520; CAC36101.1; JOINED.
 DR EMBL: AJ288521; CAC36101.1; JOINED.
 DR EMBL: AJ288522; CAC36101.1; JOINED.
 DR EMBL: AJ288523; CAC36101.1; JOINED.
 DR EMBL: AJ288524; CAC36101.1; JOINED.
 DR EMBL: AJ288525; CAC36101.1; JOINED.
 DR EMBL: AJ288526; CAC36101.1; JOINED.
 DR EMBL: AJ288527; CAC36101.1; JOINED.
 DR EMBL: AJ288528; CAC36101.1; JOINED.
 DR EMBL: AJ288529; CAC36101.1; JOINED.
 DR EMBL: AJ288530; CAC36101.1; JOINED.
 DR EMBL: AJ288531; CAC36101.1; JOINED.
 DR EMBL: AJ288532; CAC36101.1; JOINED.
 DR EMBL: AJ288533; CAC36101.1; JOINED.
 DR EMBL: AJ288534; CAC36101.1; JOINED.
 DR EMBL: AJ288535; CAC36101.1; JOINED.
 DR EMBL: AJ288536; CAC36101.1; JOINED.
 DR EMBL: AJ288537; CAC36101.1; JOINED.
 DR EMBL: AJ288538; CAC36101.1; JOINED.
 DR EMBL: M92993; AAA21610.1; -;
 DR EMBL: M55790; AAB19637.1; -;
 DR EMBL: M81379; AAB51556.1; -;
 DR EMBL: L08650; AAA52044.1; -;
 DR EMBL: U02519; AAA18942.1; -;
 Alignment Scores:
 Pred. No.: 17.3 Length: 1670
 Score: 56.50 Matches: 16
 Percent Similarity: 54.05% Conservative: 4
 Best Local Similarity: 43.24% Mismatches: 12
 Query Match: 27.16% Indels: 5
 DB: 1 Gaps: 3
 US-09-513-999c-3792_copy_51_161 (1-111) x CA34_HUMAN (1-1670)
 QY 3 GGGTGGATGCTTTGGCTGGAGATCTTTTCATCTTT-----GCAGGGACTTCT 53
 Db 1573 GlyTThrGlyAlaIleuAlaSerProGlySer---CysLeuGluGluPhe 1592
 QY 54 GGGGGCCGGA--GTATGTAAACCTCGTGTGTGTGTGCTGAGTGG 101
 Db 1593 GlyTThrGlyAlaIleuAlaSerProGlySer---CysLeuGluGluPhe 1608
 RESULT 9
 ID CATE_HUMAN STANDARD; PRT; 396 AA.
 AC P14091;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cathepsin E precursor (EC 3.4.23.34).
 GN CTSE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=89380302; PubMed=2674141;
 RT Azuma T., Pals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;
 RT "Human gastric cathepsin E. Predicted sequence, localization to
 RT chromosome 1, and sequence homology with other aspartic
 RT proteinases.";
 RL J. Biol. Chem. 264:16748-16753(1989).
 [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92112877; PubMed=1370478;
 RT Azuma T., Liu W.G., Vander laan D.J., Bowcock A.M., Taggart R.T.;
 RT "Human gastric cathepsin E gene. Multiple transcripts result from
 RT alternative polyadenylation of the primary transcripts of a single
 RT gene locus at 1q31-q32.";
 RL J. Biol. Chem. 267:1609-1614(1992).
 [3]
 RP SEQUENCE FROM N.A.
 RA Tanneil P.J., Kay J.;
 RT "Human procathepsin E.";
 RT Submitted (Oct-1999) to the EMBL/Genbank/DBJ databases.
 [4]
 RP SEQUENCE OF 54-68; 77-95; 141-155; 275-285 AND 389-396.
 RA MEDLINE=90241267; PubMed=2334440;
 RT Athauda S.B.P., Matsuzaki O., Kageyama T., Takahashi K.;
 RT "Structural evidence for two isozymic forms and the carbohydrate
 RT attachment site of human gastric cathepsin E.";
 RL Biochem. Biophys. Res. Commun. 168:878-885(1990).
 CC -1- FUNCTION: DUE OF ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
 CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
 CC -1- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
 CC specificity.
 CC -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M64424; AAA52300.1; -;
 DR EMBL: M64413; AAA52300.1; JOINED.
 DR EMBL: M64417; AAA52300.1; JOINED.
 DR EMBL: M64418; AAA52300.1; JOINED.
 DR EMBL: M64419; AAA52300.1; JOINED.
 DR EMBL: M64420; AAA52300.1; JOINED.
 DR EMBL: M64421; AAA52300.1; JOINED.
 DR EMBL: M64422; AAA52300.1; JOINED.
 DR EMBL: J05036; AAA52130.1; -;
 DR EMBL: AJ250717; CAB82850.1; -;
 DR PIR: A34401; A34401.
 DR PIR: A34643; A34643.
 DR PIR: A42038; A42038.
 DR HSSP: P00794; 4CMS.
 DR HSSP: A01_010; -;
 DR Genew: HGNC:2530; CTSE.
 DR MIM: 116890; -;
 DR InterPro: IPR001461; AsproteaseA1.
 DR InterPro: IPR001969; Asprotease_sib.
 DR Pfam: PF00026; asp_1.
 DR PRINTS: PR00792; PEPsin.
 DR PROSITE: PS00141; ASP_PROTEASE; 2.
 KM Hydrolyase; Aspartyl protease; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 17
 FT PROPEP 18 53
 FT CHAIN 54 396
 FT MOD_RSS 18 18
 FT ACT_SITE 96 96
 FT BY SIMILARITY.

FT ACT_SITE 281 281 BY SIMILARITY.
 FT DISULFID 60 60 INTERCHAIN (PROBABLE).
 FT DISULFID 109 114 BY SIMILARITY.
 FT DISULFID 272 276 BY SIMILARITY.
 FT DISULFID 314 351 BY SIMILARITY.
 FT CARBOHYD 90 90 N-LINKED (GLCNAC...).
 FT CARBOHYD 220 220 O-LINKED (POTENTIAL).
 FT CARBOHYD 333 333 O-LINKED (POTENTIAL).
 SQ SEQUENCE 396 AA; 42793 MW; 40B643C5FB01521E CRC64;
 Alignment Scores:
 Pred. NO: 20 Length: 396
 Score: 56.00 Matches: 12
 Percent Similarity: 62.07% Conservative: 6
 Best Local Similarity: 41.38% Mismatches: 11
 Query Match: 26.92% Indels: 0
 DB: Gaps: 0
 US-09-513-999c-3792_COPY_51_161 (1-111) x CATE_HUMAN (1-396)
 QY 4 GGTCGATCTTTGGCTTCAGGATCTTTTCATCTTGCAGGACTTTCGGCCGAG 63
 Db 132 GlnGlnSerPheSerIleGlnIleGlyThrGlySerLeuSerGlyIleIleGlyAlaAsp 151
 QY 64 TATGTAAACCTCGGCTCTCTGTG 90
 Db 152 GlnValSerValGlnGlyLeuThrVal 160
 RESULT 10
 TTHY_MOUSE STANDARD: PRT: 147 AA.
 AC P07309;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transhyretin precursor (Prelalbumin).
 GN TTR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=86139970; PubMed=3005251;
 RA Wakasugi S., Maeda S., Shimada K., Nakashima H., Migita S.;
 RT "Structural comparisons between mouse and human prealbumin.";
 RT J. Biochem. 98:1707-1714(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=87008480; PubMed=3020014;
 RA Wakasugi S., Maeda S., Shimada K.;
 RT "Structure and expression of the mouse prealbumin gene.";
 RT J. Biochem. 100:49-58(1986).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Choroid plexus;
 RA Kita H., Kawamoto S., Okubo K., Matsubara K.;
 RT Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: Thyroid hormone-binding protein. Probably transports
 CC thyroxine from the bloodstream to the brain.
 CC -1- SUBUNIT: Homotrimer.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN THE CHOROID PLEXUS. ALSO
 CC PRESENT IN THE LIVER.
 CC -1- SIMILARITY: BELONGS TO THE TRANSHYRETIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----


```

CC -----
DR EMBL: X03351; CA27057.1; -
DR EMBL: D00073; BAA0050.1; -
DR EMBL: D00071; BAA0050.1; JOINED.
DR EMBL: D00072; BAA0050.1; JOINED.
DR EMBL: D89076; BAA13757.1; -
DR PIR: A24132; VBMS.
DR HSSP: P02767; IGKE.
DR MGD: MGI:38865; TLR.
DR InterPro: IPR000895; Transthyretin.
DR Pfam: PF00576; Transthyretin_1.
DR PRINTS: PR00189; TRNSTHYRETIN.
DR ProDom: PD003457; Transthyretin_1.
DR SMART: SM00095; TR_THY_1.
DR PROSITE: PS00768; TRANSTHYRETIN_1; 1.
DR PROSITE: PS00769; TRANSTHYRETIN_2; 1.
DR Albumin; Transport; Retinol-binding; Vitamin A; Thyroid hormone;
KW Signal.
FT CHAIN 1 20 TRANSTHYRETIN.
FT BINDING 35 35 THYROID HORMONES (BY SIMILARITY).
FT BINDING 74 74 THYROID HORMONES (BY SIMILARITY).
SQ SEQUENCE 147 AA; 15776 MW; 9803CCC3024BA911 CRC64;

Alignment Scores:
Pred. No.: 23.2 Length: 147
Score: 55.50 Matches: 12
Percent Similarity: 64.00% Conservative: 4
Best Local Similarity: 48.00% Mismatches: 8
Query Match: 26.68% Indels: 1
DB: 1 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x TTHY_MOUSE (1-147)

OY 6 TGGATCTTTGCTTGCAGAGATCTTTTCATCTTT---GCAGGAGATCTTGGGCGCGGA 62
DB 7 PheLeuLeuCySLeuAlaGlyLeuValPheValSerGluAlaGlyProAlaGlyValaGly 26
OY 63 GATGTAAACTCTCT 77
DB 27 GluSerLysCysPro 31

RESULT 11
VS41_GIALA STANDARD: PRT: 687 AA.
AC P92127;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE Variant-specific surface protein VSP41 precursor (CRISP-90).
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O2-4A1;
RX MEDLINE=97321554; PubMed=9178264;
RA Papanastasiou P., Bruderer T., Li Y., Bommeili C., Koehler P.;
RT "Primary structure and biochemical properties of a variant-specific
RT surface protein of Giardia."
RL Mol. Biochem. Parasitol. 86:13-27(1997).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=97233006; PubMed=9078242;
RA Papanastasiou P., McConville M.J., Ralton J., Koehler P.;
RT "The variant-specific surface protein of Giardia, VSP41, is a
RT glycosylated and palmitoylated protein."
RL Biochem. J. 332:49-56(1997).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
CC PLASMA MEMBRANE.
CC -1- PTM: O-GLYCOSYLATED. THE MAJOR GLYCAN IS A TRISACCHARIDE WITH GLC
CC AT THE REDUCING TERMINUS.
CC -1- PTM: PALMITOYLATED.

```

```

CC -----
CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
CC -----
DR EMBL: Z83743; CAB06038.1; -
DR HSSP: O14763; IDOG.
DR GlycoSuiteDB: P92127; -
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR005127; Giardia_VSP.
DR Pfam: PF03302; VSP; 2.
DR SMART: SM00181; EGF; 2.
DR SMART: SM00001; EGF_Like; 1.
DR SMART: SM00261; FU_3.
KW Antigen; Repeat; Transmembrane; Glycoprotein; Lipoprotein; Palmitate;
KW Signal.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 687 VARIANT-SPECIFIC SURFACE PROTEIN VSP41.
FT DOMAIN 15 660 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 661 681 POTENTIAL.
FT DOMAIN 682 687 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 687 AA; 70857 MW; D892F675D626D7EC CRC64;

Alignment Scores:
Pred. No.: 23.3 Length: 687
Score: 55.50 Matches: 12
Percent Similarity: 41.03% Conservative: 4
Best Local Similarity: 30.77% Mismatches: 8
Query Match: 26.68% Indels: 15
DB: 1 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x VS41_GIALA (1-687)

OY 12 TTTTGCTTGCAGAGATCTTTTCATCTTTGCAGG-----
DB 500 TyrCysAlaAlaGlyPheLeuTyrMetGlyGlyCysTyrLysIleAspThrValPro 519
OY 48 -----ACCTTGGGGCGGAGTAGTAACCTCGGCTC 83
DB 520 GlySerTyrMetCysSerLysSerThrThrAlaGlyValCysAspThrProAsnAla 538

RESULT 12
N12B_MEDSA STANDARD: PRT: 113 AA.
AC Q40339; Q40342;
DT 15-DEC-1998 (Rel. 37; Created)
DT 15-DEC-1998 (Rel. 37; Last sequence update)
DT 15-DEC-1998 (Rel. 37; Last annotation update)
DE Early nodulin 12B precursor (N-12B).
GN ENOD12B.
OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
OX NCBI_TaxID=3879;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nagyszénasi;
RX MEDLINE=93144711; PubMed=7678770;
RA Allison L.A., Kiss G.B., Bauer P., Poliet M., Pierre M.,
RA Savoure A., Kondorosi E., Kondorosi A.;
RT "Identification of two alfalfa early nodulin genes with homology to
RT members of the pea Enod12 gene family."
RL Plant Mol. Biol. 21:375-380(1993).
RN [2]
RP SEQUENCE OF 15-113 FROM N.A.
RC STRAIN=cv. COERULA W2;

```

```

RT other properties.";
RN Eur. J. Biochem. 230:977-986(1995).
RL [2]
RP SEQUENCE OF 4-150 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97083734; PubMed=6930078;
RA Archibald A.L., Couperwhite S., Jiang Z.H.;
RT "the porcine TTR locus maps to chromosome 6q.";
RL Ann. Genet. 27:351-353(1996).
CC -1- FUNCTION: Thyroid hormone-binding protein. Probably transports
CC thyroxine from the bloodstream to the brain.
CC -1- SUBUNIT: Homotrimer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SYNTHESIZED AND SECRETED BY THE CHOROID
CC PLEXUS. IT IS ALSO PRESENT IN THE LIVER OF BIRDS AND EUTHERIANS.
CC -1- SIMILARITY: BELONGS TO THE TRANSTHYRETIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X82258; CAA57713.1; -
DR EMBL; U16131; AAA79042.1; -
DR EMBL; X87846; CAA61120.1; -
DR HSSP; P02767; IGRF
DR InterPro; IPR000895; Transthyretin.
DR Pfam; PF00576; Transthyretin; 1.
DR PRINTS; PR00189; TRANSTHYRETIN.
DR ProDom; PD003457; Transthyretin; 1.
DR SMART; SM00095; TR_THY. 1.
DR PROSITE; PS00768; TRANSTHYRETIN_1; 1.
DR PROSITE; PS00769; TRANSTHYRETIN_2; 1.
KW Albumin; Transport; Retinol-binding; Vitamin A; Thyroid hormone;
KW Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 1 150 TRANSTHYRETIN.
FT BINDING 36 36 THYROID HORMONES (BY SIMILARITY).
FT FT 74 74 THYROID HORMONES (BY SIMILARITY).
SQ SEQUENCE 150 AA; 16081 MW; 0427ED5D3094CA07 CRC64;

Alignment Scores:
Pred. NO.: 31.4 Length: 150
Score: 54.50 Matches: 12
Percent Similarity: 62.50% Conservative: 3
Best local Similarity: 50.00% Mismatches: 8
Query Match: 26.20% Indels: 1
DB: 1 Gaps: 1

US-09-513-999C-3792_COPY_51_161 (1-111) x TTHY_PIG (1-150)
QY 9 ATCTTTTGCCCTTGACGATCTTTTCATCTT---GCAGGACCTTGCGGCGGAGTA 65
::: ||||| ||||| ||||| ||||| |||||
Db 8 leuLeucycysleuAlaaglyleuValAlpheValsercIuAlaaglyProAlaaglyGlu 27
QY 66 TGTAAACTCT 77
||| |||
Db 28 serIyscysPro 31

RESULT 14
UT2_RAT
AC 062668; STANDARD; PRT; 397 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE urea transporter, kidney.
GN SIC14A2 OR UT2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

DT 01-MAY-1992 (Rel. 22, Last Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Glycoprotein polypeptide precursor [Contains: glycoprotein G1;
GN glycoprotein G2].
GN GPC.
OS Junin arenavirus.
OC Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.
OX NCBI_TaxID=11619;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC2;
RC MEDLINE=91374010; PubMed=1654373;
RA Chillinghelli P.D., Rivera-Pomar R.V., Lozano M.E., Grau O.,
RA Romanowski V.;
RT "Molecular organization of Junin virus S RNA: complete nucleotide
RT sequence, relationship with other members of the Arenaviridae and
RT unusual secondary structures.;"
RL J. Gen. Virol. 72:2129-2141(1991).
-1- SIMILARITY: BELONGS TO THE ARENAVIRUSES GPC PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL, D10072; BAA00964.1; -.
DR PIR, J00978; VGXPJV.
DR InterPro: IPR001535; Arena.glycoprot.
DR Pfam: PF00798; Arena.glycoprot; 1.
KW Polypeptide; Glycoprotein; Envelope protein.
KW CHAIN 1 244 GLYCOPROTEIN G1.
FT CAROBYD 91 481 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROBYD 101 101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROBYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROBYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROBYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROBYD 361 361 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROBYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CAROBYD 383 383 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 481 AA; 55118 MW; 6E3BE4E587CE27A3 CRC64;

Alignment Scores:
Pred. No.: 31.5 Length: 481
Score: 54.50 Matches: 11
Percent Similarity: 61.54% Conservative: 5
Best Local Similarity: 42.31% Mismatches: 9
Query Match: 26.20% Indels: 1
DB: 1 Gaps: 1

US-09-513-999C-3792_COPY_51_161 (1-111) x VGLY-JUNIN (1-481)
QY 30 TTTTCATCTTTCGACGAGGACTTCGGGGCGAGTAGTAAACTCGGAGCTCTGTGT 89
Db 249 PhepSerTrpSerLeuThrAspSerSerGlyLys--AspTrpProGlyGlyTyrCys 267
QY 90 GTGCTGAGTGAGCTGCTC 107
Db 268 LeuGluGluTrpMetLeu 273

Search completed: April 30, 2003, 14:35:11
Job time : 12 secs

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 30, 2003, 14:34:49 ; Search time 14.5 Seconds

1471.851 Million cell updates/sec

Title: US-09-513-999C-3792_COPY_51_161

Perfect score: 208
Sequence: 1 atggtgatcttctgctt.....gcctagtgctgctctact 111

| Scoring table: | | BLOSUM62 |
|----------------|------|-------------|
| Xgapop | 10.0 | Xgapext 0.5 |
| Ygapop | 10.0 | Ygapext 0.5 |
| Fgapop | 6.0 | Fgapext 7.0 |
| Delop | 6.0 | Delext 7.0 |

searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

Command line parameters:

```

-Model=Iramet+n2p.model -DEV=fast1
-0/cgna21/USPTO.spool/US09513999/runat.30042003.143402.25617/app.query.fasta.1.2633
-DB=PIR.73 -GMMT=fastan -SUFFIX=12p.rpr -MINMATCH=0.1 -LOOPEXT=0
-DONTS-bits -START=1 -END=1 -MATRIX=dl0sum62 -TRANS-human.cdi -LIST=45
-DUALALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09513999 -CGEN 1 1 62 -ATTEN 30042003.143402.25617 -NCP=6 -ICPU=3
-NO_XLXPY -NO_MMP -IAREJOURY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMECOUT=120
-WARN_TIMECOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database :

```
1:  pir1:*
2:  pir2:*
3:  pir3:*
4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|--------------------|
| C 1 | 63.5 | 31.4 | 634 | 2 | T47156 | hypothetical prote |
| C 2 | 60.5 | 29.1 | 471 | 2 | A39024 | collagen alpha 3(I |
| C 3 | 60 | 28.8 | 314 | 2 | T16300 | hypothetical prot |
| C 4 | 60 | 28.8 | 1145 | 2 | T51546 | cellulose synthas |
| C 5 | 60 | 29.7 | 5170 | 2 | T15348 | hypothetical prote |
| C 6 | 59 | 28.4 | 382 | 2 | B97583 | hypothetical prot |
| C 7 | 59 | 28.4 | 382 | 2 | AB2804 | hypothetical prot |
| C 8 | 59 | 29.2 | 1297 | 2 | S25712 | son-of-sevenless-7 |
| C 9 | 58 | 28.7 | 197 | 2 | T04782 | hypothetical prot |
| C 10 | 58 | 27.9 | 391 | 2 | A43356 | cathpsin E (EC 3. |
| C 11 | 57 | 27.4 | 240 | 2 | T12191 | NADH dehydrogen |
| C 12 | 57 | 28.2 | 880 | 2 | T04523 | hypothetical prot |
| C 13 | 57 | 28.2 | 3149 | 1 | Q00B5 | BLP1 protein - hu |
| C 14 | 56.5 | 28.0 | 938 | 2 | A56731 | chromatin assemb |

| | | | | | | |
|----|------|------|------|---|--------|--------------------|
| 15 | 56.5 | 27.2 | 1670 | 1 | CGH43B | collagen alpha 3(I |
| 16 | 56 | 26.9 | 335 | 1 | S74441 | iron(III) dicitrat |
| 17 | 56 | 26.9 | 395 | 2 | A34401 | cathpsin E (EC 3. |
| 18 | 55.5 | 26.7 | 147 | 1 | VBMS | transhyretin prec |
| 19 | 55.5 | 27.5 | 504 | 2 | T13475 | transhyretin prote |
| 20 | 55.5 | 27.5 | 725 | 2 | D81976 | probable ferric si |
| 21 | 55.5 | 27.5 | 725 | 2 | H81030 | TonB-dependent rec |
| 22 | 55 | 27.2 | 239 | 2 | A70626 | hypothetical prote |
| 23 | 55 | 26.4 | 323 | 2 | B95888 | probable dipeptida |
| 24 | 55 | 27.2 | 342 | 2 | T39545 | hypothetical prote |
| 25 | 55 | 26.4 | 830 | 2 | T01058 | hypothetical prote |
| 26 | 54.5 | 27.0 | 113 | 2 | S31076 | ENOD12B protein pr |
| 27 | 54.5 | 26.2 | 150 | 2 | S65955 | transhyretin prec |
| 28 | 54.5 | 26.2 | 481 | 1 | VGXBJV | surface glycoprote |
| 29 | 54.5 | 27.0 | 816 | 2 | C88196 | protein ZK127.7 |
| 30 | 54.5 | 26.2 | 835 | 2 | T39106 | vacuolar protein s |
| 31 | 54.5 | 27.0 | 1520 | 2 | T23620 | hypothetical prote |
| 32 | 54 | 26.0 | 98 | 2 | T37294 | r-cell receptor de |
| 33 | 54 | 26.7 | 150 | 2 | E72587 | hypothetical prote |
| 34 | 54 | 26.0 | 257 | 2 | S55612 | hypothetical prote |
| 35 | 54 | 26.0 | 387 | 1 | A48326 | ubiquinol-cytochro |
| 36 | 54 | 26.7 | 388 | 2 | A12539 | hypothetical prote |
| 37 | 54 | 26.7 | 706 | 2 | T25897 | hypothetical prote |
| 38 | 53.5 | 25.7 | 246 | 2 | I48302 | collagen alpha 3(I |
| 39 | 53.5 | 26.5 | 384 | 2 | H87019 | probable glycosyl |
| 40 | 53.5 | 26.5 | 534 | 2 | E86170 | hypothetical prote |
| 41 | 53.5 | 26.5 | 634 | 1 | S35574 | transcription fact |
| 42 | 53.5 | 26.5 | 776 | 2 | T19900 | hypothetical prote |
| 43 | 53 | 26.2 | 118 | 1 | M4W13 | E4 protein - human |
| 44 | 53 | 26.2 | 346 | 2 | H70195 | saloglycoproteina |
| 45 | 53 | 26.2 | 361 | 2 | T42525 | splicing factor-11 |

ALIGNMENTS

RESULT 1

hypothetical protein DKFzp727E211.1 - human (fragment)

```
C:\species: Homo sapiens (man)
C:\Date: 20-Apr-2000 #sequence
```

C;Accession: T47156
P.Dowdka A : Wellerauther P : Moves H W : Weil B : Wiemann S

submitted to the Protein Sequence Database, March 2000

A;Accession: T47156

A;Molecule type: mRNA

A;Residues: 1-634 <AA

A; Experimental source

A; Note: DKFZp727E211.

Alignment Scores:

Pred. No.:

Percent Similarity:

Query Match:

DB: 2 caps. 3

HS-09-513-999C-3792 COPY 51 161 (1-111) X T47156 (1-634)

[illegible][illegible]

| Country | Year | Population (millions) | GDP (billion USD) | Urban population (%) | Life expectancy (years) | Infant mortality (per 1,000 live births) | Healthcare expenditure (USD per capita) |
|-----------------------------|------|-----------------------|-------------------|----------------------|-------------------------|--|---|
| USA | 2000 | 280 | 10,000 | 75 | 77 | 12 | 1,200 |
| China | 2000 | 1,200 | 10,000 | 35 | 72 | 25 | 100 |
| India | 2000 | 1,000 | 1,000 | 25 | 65 | 40 | 50 |
| Japan | 2000 | 125 | 5,000 | 92 | 82 | 8 | 1,500 |
| Germany | 2000 | 82 | 3,000 | 88 | 78 | 7 | 1,000 |
| France | 2000 | 64 | 2,500 | 85 | 79 | 6 | 1,100 |
| UK | 2000 | 60 | 2,000 | 83 | 78 | 6 | 1,000 |
| Italy | 2000 | 59 | 2,000 | 80 | 79 | 6 | 1,000 |
| Spain | 2000 | 45 | 1,500 | 78 | 81 | 5 | 1,000 |
| Sweden | 2000 | 9 | 300 | 88 | 81 | 4 | 1,200 |
| Norway | 2000 | 4 | 200 | 85 | 80 | 3 | 1,300 |
| Denmark | 2000 | 5 | 200 | 85 | 80 | 3 | 1,200 |
| Netherlands | 2000 | 16 | 400 | 85 | 80 | 3 | 1,100 |
| Belgium | 2000 | 10 | 300 | 85 | 80 | 3 | 1,000 |
| Austria | 2000 | 8 | 200 | 85 | 80 | 3 | 1,000 |
| Switzerland | 2000 | 7 | 300 | 85 | 80 | 3 | 1,200 |
| South Korea | 2000 | 45 | 1,000 | 85 | 75 | 10 | 500 |
| Singapore | 2000 | 4 | 100 | 100 | 75 | 10 | 500 |
| Hong Kong | 2000 | 6 | 200 | 100 | 75 | 10 | 500 |
| Taiwan | 2000 | 22 | 500 | 85 | 75 | 10 | 500 |
| South Africa | 2000 | 40 | 100 | 55 | 52 | 30 | 100 |
| Brazil | 2000 | 170 | 1,000 | 75 | 65 | 20 | 100 |
| Mexico | 2000 | 100 | 1,000 | 70 | 65 | 20 | 100 |
| Argentina | 2000 | 40 | 1,000 | 80 | 75 | 15 | 200 |
| Chile | 2000 | 16 | 200 | 80 | 75 | 15 | 200 |
| Colombia | 2000 | 40 | 100 | 60 | 65 | 25 | 100 |
| Venezuela | 2000 | 26 | 200 | 60 | 65 | 25 | 100 |
| Ecuador | 2000 | 12 | 100 | 60 | 65 | 25 | 100 |
| Peru | 2000 | 27 | 100 | 60 | 65 | 25 | 100 |
| Bolivia | 2000 | 9 | 100 | 60 | 65 | 25 | 100 |
| Paraguay | 2000 | 7 | 100 | 60 | 65 | 25 | 100 |
| Uruguay | 2000 | 3.5 | 100 | 60 | 65 | 25 | 100 |
| Costa Rica | 2000 | 4 | 100 | 60 | 65 | 25 | 100 |
| Panama | 2000 | 3 | 100 | 60 | 65 | 25 | 100 |
| Dominican Republic | 2000 | 7 | 100 | 60 | 65 | 25 | 100 |
| Jamaica | 2000 | 2.7 | 100 | 60 | 65 | 25 | 100 |
| Trinidad and Tobago | 2000 | 1.2 | 100 | 60 | 65 | 25 | 100 |
| Barbados | 2000 | 0.3 | 100 | 60 | 65 | 25 | 100 |
| Suriname | 2000 | 0.5 | 100 | 60 | 65 | 25 | 100 |
| Guyana | 2000 | 0.8 | 100 | 60 | 65 | 25 | 100 |
| French Guiana | 2000 | 0.2 | 100 | 60 | 65 | 25 | 100 |
| Guadeloupe | 2000 | 0.1 | 100 | 60 | 65 | 25 | 100 |
| Martinique | 2000 | 0.1 | 100 | 60 | 65 | 25 | 100 |
| Reunion | 2000 | 0.8 | 100 | 60 | 65 | 25 | 100 |
| Mayotte | 2000 | 0.1 | 100 | 60 | 65 | 25 | 100 |
| French Polynesia | 2000 | 0.2 | 100 | 60 | 65 | 25 | 100 |
| New Caledonia | 2000 | 0.2 | 100 | 60 | 65 | 25 | 100 |
| Wallis and Futuna | 2000 | 0.05 | 100 | 60 | 65 | 25 | 100 |
| French Southern Territories | 2000 | 0.01 | 100 | 60 | 65 | 25 | 100 |

56 CCAGAGTCCCTGCATGATGAAATGMAI CCGAAGGCAATGAT

Db 305 AspArgSer-----LysAspGluGluGluCysLeuGluGluGluAsp 31

RESULT

collagen alpha 3(IV) chain - bovine (fragment)
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 04-Dec-1992 #sequence, revision 04-Dec-1992 #text, change 13-Aug-1999
 C/Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
 R/Morrison, K.E.; Germino, G.G.; Reeder, S.T.
 J. Biol. Chem. 266, 34-39, 1991
 A/Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the
 A/Reference number: A39024; MUID:91093146; PMID:1985905
 A/Accession: A39024
 A/Molecule type: mRNA
 A/Residues: 1-471 <MOR>
 A/Cross-references: EMBL:M63139; NID:g162886; PID:AA62708.1; PID:g162887
 R/Butkowski, R.J.; Langewald, J.P.M.; Wiestlander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 262, 7874-7877, 1987
 A/Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
 A/Reference number: S18432; MUID:87222419; PMID:2438283
 A/Accession: S20672
 A/Molecule type: Protein
 A/Residues: 227-228, 'X', 230-244 <BUT>
 C/Saus, J.; Wiestlander, J.; Langewald, J.P.M.; Quinones, S.; Hudson, B.G.
 J. Biol. Chem. 263, 1374-1380, 1988
 A/Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen
 A/Reference number: S17802; MUID:88330844; PMID:3417661
 A/Accession: S17802
 A/Molecule type: Protein
 A/Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
 R/Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
 J. Biol. Chem. 265, 5466-5469, 1990
 A/Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type
 A/Reference number: A35167; MUID:90202779; PMID:2318822
 A/Accession: A35167
 A/Molecule type: Protein
 A/Residues: 236-258 <GUN>
 R/Gunwar, S.; Ballesster, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe
 J. Biol. Chem. 266, 15318-15324, 1991
 A/Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll
 A/Reference number: A39419; MUID:91332055; PMID:1869555
 A/Accession: C39419
 A/Molecule type: Protein
 A/Residues: 236-255 <GUN>
 C/Superfamily: collagen alpha 1(IV) chain
 C/Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
 F:1-238/Domain: collagenous (fragment) #status predicted <COL>
 F:239-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
 F:239-353/Domain: repeat NC1 #status predicted <NC1>
 F:354-471/Domain: repeat NC1 #status predicted <NC1>
 F:232,238/Modified site: hydroxyproline (Pro) #status experimental
 F:306-312,417-423/Disulfide bonds: #status predicted

Alignment Scores:
 Pred. No.: 9.67 Length: 471
 Score: 60.50 Matches: 17
 Percent Similarity: 56.76% Conservative: 4
 Best Local Similarity: 45.95% Mismatches: 11
 Query Match: 29.09% Indels: 5
 DB: 2 Gaps: 3

US-09-513-999c-3792_COPY_51_161 (1-111) x A39024 (1-471)

OY 3 GGGTGATCTTTGCGAGATCTTTTCATCTT-----GCGAGGACTTCT 53
 |||||
 Db 374 GYTRPLRLEUTRPLYSGLYPHESRPHLEMERHETHRSERIAGLYSERGL 393
 |||||
 OY 54 GGGGCCGA---GTATGTAACCTCTGGGTCTGTGTGCTGCTAGTGG 101
 |||||
 Db 394 GYLAELALYGLNALALALALASERPROGLYSER---CYSLEUGLUGLUPHE 409
 |||||
 RESULT 3
 T16300
 hypothetical protein F38B6.7 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 20-Sep-1999 #sequence, revision 20-Sep-1999 #text, change 20-Sep-1999
 C/Accession: T16300

R:Wilson, L.
 submitted to the EMBL Data Library, November 1995
 A/Description: The sequence of C. elegans cosmid F38B6.
 A/Reference number: Z18491
 A/Accession: T16300
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-314 <WIL>
 A/Cross-references: EMBL:U40060; NID:g1055174; PID:g1055179; PID:AAA81145.1; CESP:F3
 A/Genes: CESP:F38B6.7
 A/Introns: 44/1; 120/1; 179/1; 219/3; 258/3

Alignment Scores:
 Pred. No.: 11.5 Length: 314
 Score: 60.00 Matches: 11
 Percent Similarity: 62.96% Conservative: 6
 Best Local Similarity: 40.74% Mismatches: 6
 Query Match: 28.85% Indels: 1
 DB: 2 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x T16300 (1-314)

OY 11 CTTTTCCTTTCGAGATCTTTTCATCTT-----TGCAGGACTTCTGGGCG 58
 |||||
 Db 278 ILEULYSCYSGLYMETILEPHEMETLEULIETHRILEUHLISLEUARGSPHETRPLY 297
 |||||
 OY 59 CGAGATGTAAACCTCTGG 79
 |||||
 Db 298 ARGLYSLEUGLUNASNPTR 304
 |||||

RESULT 4
 T15346
 cellulose synthase catalytic subunit-like protein - Arabidopsis thaliana
 N/Alternate names: protein F2K13.60
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 18-Aug-2000 #sequence, revision 18-Aug-2000 #text, change 18-Aug-2000
 C/Accession: T15346
 R/Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.;
 submitted to the Protein Sequence Database, August 2000
 A/Reference number: Z25394
 A/Accession: T15346
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1145 <SAT>
 A/Cross-references: EMBL:AL391141
 A/Experimental source: cultivar Columbia; BAC clone F2K13
 C/Genetics:
 A/Map position: 5
 A/Introns: 297/2; 566/3
 A/Notes: F2K13_60

Alignment Scores:
 Pred. No.: 10.7 Length: 1145
 Score: 60.00 Matches: 13
 Percent Similarity: 57.14% Conservative: 3
 Best Local Similarity: 46.43% Mismatches: 11
 Query Match: 28.85% Indels: 1
 DB: 2 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x T15346 (1-1145)

OY 8 GATCTTTGCGAGATCTTTTCATCTTTCGAGGACTTCTGGGCGGAGATG 67
 |||||
 Db 145 ASPLLEUPLROCYGLUCYSAPHELYSLIYCYSATGAPCYSPHE-ILEAPALAVA 164
 |||||
 OY 68 TAAACTCTGGGTCTGTGT 89
 |||||
 Db 164 ILYSTRGLYGLYGLYILEYCS 171
 |||||

RESULT 5
 T15348
 hypothetical protein B0350.1 - Caenorhabditis elegans

0
1
2
3
4
5
6
7
8
9

Score:
Percent:

File: NO.:
Score:

Best Local Simila

Query Match: 29.21% Indels: 6
DB: 2 Gaps: 1

US-09-513-999c-3792_COPY_51_161 (1-111) x S25714 (1-1297)

QY 106 AGCAGCCTCAGAGCAGACAGACCAGAGTTCATATCTCCGGCCAGAGATC 47
|||||
DB 1244 SerLeuHisSer-----SerLeuAlaHisLeuProLarProVal 1257
|||
QY 46 CCTGCAGAGTGAAGATCCTGCAGAGCAAGATCCACCCA 2
|||
DB 1258 ProProArgGlnAsnSerSerProLeuProLysLeuProPro 1272

RESULT 9

T04782
hypothetical protein F10M10.130 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999

C:Accession: T04782

R:Bevan, M.; De Haan, M.; Maarse, A.C.; Grivell, L.A.; Hohelsel, J.; Mewes, H.W.; Mayer, submitted to the Protein Sequence Database, February 1999

A:Reference number: Z15384

A:Accession: T04782

A:Molecule type: DNA

A:Residues: 1-197 <BEV>

A:Cross-references: EMBL:AL035521

A:Experimental source: cultivar Columbia; BAC clone F10M10

C:Genetics:

A:Map position: 4

A:Introns: 49/1; 76/3; 121/3; 156/3

A>Note: F10M10.130

Alignment Scores:

Pred. No.: 21.9 Length: 197
Score: 58.00 Matches: 14
Percent Similarity: 60.87% Conservative: 0
Best Local Similarity: 60.87% Mismatches: 8
Query Match: 28.71% Indels: 1
DB: 2 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x T04782 (1-197)

QY 72 TTTTACTACTCCGCCAGAGTCCCTGCAAGATGAAAAAGATCTCGACAGC 13
|||||
DB 151 PherylleuArgLysGlyLysArg-CysAsnAspGluLysGluAspLysCysAs 170
|||
QY 12 AGATCCA 6
|||||
DB 170 naspPro 172

RESULT 10

A43356
cathepsin E (EC 3.4.23.34) precursor - guinea pig

N:Alternate names: erythrocyte membrane aspartic proteinase; slow-moving proteinase
C:Species: Cavia porcellus (guinea pig)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999

C:Accession: A43356

R:Kageyama, T.; Ichinose, M.; Tsukada, S.; Miki, K.; Kurokawa, K.; Koiwai, O.; Tanji, M.

J. Biol. Chem. 267, 16450-16459, 1992

A:Title: Gastric procathepsin E and procathepsin from guinea pig. Purification, molecu

A:Reference number: A43356; MUID:92355614; PMID:1644829

A:Accession: A43356

A:Molecule type: mRNA

A:Residues: 1-391 <KAG>

A:Cross-references: GB:M8653; NID:q191294; PIDN:AAA37052.1; PID:q191295

A>Note: sequence extracted from NCBI backbone (NCBIR.110763, NCBIR.110769)

C:Superfamily: pepsin

C:Keywords: aspartic proteinase; hydrolase; membrane protein

Alignment Scores:

Pred. No.: 21.1 Length: 391
Score: 58.00 Matches: 12
Percent Similarity: 63.33% Conservative: 7

Best Local Similarity: 40.00% Mismatches: 11
Query Match: 27.88% Indels: 0
DB: 2 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x A43356 (1-391)

QY 1 ATGGGTGATCTTTTGGCTTCGACAGATCTTTTCATCTTTGACAGGACCTTCGGGCC 60
|||||
DB 127 ValGlyAsnSerPheSerIleGlnTrpGlyThrGlySerLeuThrGlyIleGlyAla 146
|||
QY 61 GAGTATGTAACCTCCGGGCTCTGTG 90
|||
DB 147 AspGlnValSerValGluGlyLeuThrVal 156

RESULT 11

T12191
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - Simulium vittatum mitochondr

C:Species: mitochondrion Simulium vittatum
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002

C:Accession: T12191

R:Zhu, X.; Pruess, K.P.; Powers, T.O.
submitted to the EMBL Data Library, September 1997

A:Description: Mitochondrial DNA polymorphism in a black fly, Simulium vittatum (Dipt

A:Reference number: Z17447

A:Accession: T12191

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-240 <ZHU>

A:Cross-references: EMBL:AF023535; NID:q2547181; PID:q2547182; PIDN:AAHB1263.1

A:Experimental source: strain cytospecies IIL-1 Nebraska

C:Genetics:

A:Genome: mitochondrion

C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5

C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylatio

Alignment Scores:

Pred. No.: 29.5 Length: 240
Score: 57.00 Matches: 15
Percent Similarity: 52.94% Conservative: 3
Best Local Similarity: 44.12% Mismatches: 16
Query Match: 27.40% Indels: 0
DB: 2 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x T12191 (1-240)

QY 1 ATGGGTGATCTTTTGGCTTCGACAGATCTTTTCATCTTTGACAGGACCTTCGGGCC 60
|||||
DB 196 IleGlyGlySerPheLeuSerTrpLeuMetPheProSerValSerMetIleCysLeuPro 215
|||
QY 61 GAGTATGTAACCTCCGGGCTCTGTG 102
|||
DB 216 MetTyrTyrLysAsnLeuThrLeuPheValCysLeuGlyGly 229

RESULT 12

T04523
hypothetical protein F16A16.130 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 30-Apr-1999

C:Accession: T04523

R:Bevan, M.; Brandt, P.; Dose, S.; Jayke, D.; Scharfe, M.; Schon, O.; Hohelsel, J.; M

submitted to the Protein Sequence Database, February 1999

A:Reference number: Z15376

A:Accession: T04523

A:Molecule type: DNA

A:Residues: 1-880 <BEV>

A:Cross-references: EMBL:AL035353

A:Experimental source: cultivar Columbia; BAC clone F16A16

C:Genetics:

A:Map position: 4

A:Introns: 660/1

A>Note: F16A16.130

Alignment Scores:

Pred. No.: 27.4
Score: 57.00
Percent Similarity: 67.86%
Best Local Similarity: 35.71%
Query Match: 28.22%
DB: 2
Length: 880
Matches: 10
Conservative: 9
Mismatches: 9
Indels: 0
Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x T04523 (1-880)

OY 90 CACACAGACCCAGGAGTTTTCATACCTACCGCCGACCAAGATGATAA 31

DB 189 HissertlunAlaYsargIletrValleuArgProserIylsAlaYcIunhrluYs 208

OY 30 AGAATCTGCAGGCAAAAGATCC 7

DB 209 TyrValValGlnGlyArgArgAsn 216

RESULT 13

QBRL1 protein - human herpesvirus 4 (strain B95-8)

C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 16-Jul-1999

C:Accession: G93065; A03747; S32993

R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus

A:Reference number: A93065; M01D:85035713; PMID:6092825

A:Accession: G93065

A:Molecule type: DNA

A:Residues: 1-3149 <BAN>

A:Cross-references: EMBL:V01555; NID:959074; PIDN:CAA24839.1; PID:91334853

R:Bankier, A.T.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H

Nature 310, 207-211, 1984

A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A:Reference number: A03794; M01D:84270667; PMID:6087149

A:Contents: annotation: protein coding region

C:Superfamily: human herpesvirus 4 BRL1 protein

Alignment Scores:

Pred. No.: 25.5 Length: 3149

Score: 57.00 Matches: 11

Percent Similarity: 41.18% Conservative: 3

Best Local Similarity: 32.35% Mismatches: 20

Query Match: 28.22% Indels: 0

DB: 1 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x QBRL1 (1-3149)

OY 103 AGCCACCTCAGGACACACAGACCCGAGGTTTTCATCTCCGCCCCAGATCCCT 44

DB 449 ThrHisValProProHisArgProSerAlaAlaArgLeuProProValIlePro 468

OY 43 GCAAGATGAAAGAAATCGTCGACGCAAGCAAAAGATCCACCA 2

DB 469 IleProHisGlnSerProProAlaSerProThrProHisPro 482

Pred. No.: 31.8
Score: 56.50
Percent Similarity: 44.64%
Best Local Similarity: 28.57%
Query Match: 27.97%
DB: 2
Length: 938
Matches: 16
Conservative: 9
Mismatches: 10
Indels: 21
Gaps: 4

US-09-513-999c-3792_COPY_51_161 (1-111) x A56731 (1-938)

OY 106 AGCAGCCTCAGGACACACAGACCCGAGGATTTTTCATACCTCCGCCCA 53

DB 749 SerAsnHisThrIleProserIylsAlaYcIunhrluYs 768

OY 52 GAA-----GTCCCTGCAGG-----ATGAAAAGATCT 23

DB 769 SerGlnAspAlaAlaIleProserIylsSerArgLeuIleSerGlnAsnSer 788

OY 22 GCAAGGCAAAAG-----ATCCACCA 2

DB 789 ValTyrGlnuYsArgProAspArgMetCysTrpTyrValHisPro 804

RESULT 15

CGH03B

collagen alpha 3(IV) chain precursor, long splice form - human

N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice for

C:Species: Homo sapiens (man)

C:Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999

C:Accession: A54763; A43928; A44043; A45971; A39786

R:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Trygvason, K.; Reiders, S.T.

J. Biol. Chem. 269, 23013-23017, 1994

A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpress

A:Reference number: A54763; M01D:94364994; PMID:8083201

A:Accession: A54763

A:Molecule type: mRNA

A:Residues: 1-1670 <MAR>

A:Cross-references: GB:180031; NID:9577563; PID:9577564

A:Experimental source: kidney

R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.

J. Clin. Invest. 89, 592-601, 1992

A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the

A:Reference number: A43928; M01D:92147878; PMID:1737849

A:Accession: A43928

A:Molecule type: mRNA

A:Residues: 1331-1524, 'T', 1526-1670 <TUR>

A:Cross-references: GB:M81379

A:Experimental source: kidney

R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 267, 19780-19784, 1992

A:Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpas

tion.

A:Reference number: A44043; M01D:93015826; PMID:1400291

A:Accession: A44043

A:Molecule type: DNA; mRNA

A:Residues: 1386-1670 <OUT>

A:Cross-references: GB:M92993; NID:9177895; PIDN:AAA21610.1; PID:9177896

A:Note: Sequence extracted from NCBI backbone (NCBIP:115597)

R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.

J. Biol. Chem. 269, 17358, 1994

A:Reference number: A44738; M01D:94274734; PMID:8006044

A:Contents: annotation; erratum; correction to intronic sequence in A44043

A:Reference number: A44738; M01D:94274734; PMID:8006044

A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.

A:Reference number: A45971; M01D:93280184; PMID:8505332

A:Accession: A45971

A:Molecule type: mRNA

A:Residues: 1427-1444 <BER>

A:Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly id

R:Morrisson, K.E.; Maruyama, M.; Yang-Feng, T.L.; Reiders, S.T.

Am. J. Hum. Genet. 49, 545-554, 1991

A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain

A:Reference number: A39786; M01D:91335370; PMID:1882840

A:Accession: A39786

A:Molecule type: mRNA

A:Residues: 1-938 <KXU>

A:Cross-references: GB:U20979; NID:9882257; PIDN:AAA6736.1; PID:9882258

A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain

A:Reference number: A39786; M01D:91335370; PMID:1882840

A:Accession: A39786

A:Molecule type: mRNA

A:Residues: 1-938 <KXU>

A:Cross-references: GB:U20979; NID:9882257; PIDN:AAA6736.1; PID:9882258

A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain

A:Reference number: A39786; M01D:91335370; PMID:1882840

A:Accession: A39786

A:Molecule type: mRNA

```

A:Accession: A39786
A:Molecule type: mRNA
A:Residues: 1453-1593, /5',1595-1670 <MOR>
A:Cross-references: GB:SS5790; NID:g234418; PID:AB119637.1; PID:g234419
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
C:Genetics:
A:Gene: GDB:COL4A3
A:Cross-references: GDB:128351; OMIM:120070
A:Map position: 2q36-q27
A:Introns: 1385/1, 1418/1, 1488/1, 1547/2, 1585/3, 1643/2 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-
C:Function:
A:Description: minor structural component of extracellular basement membrane in kidney
A:Superfamily: collagen alpha 1(IV) chain
A:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MNT>
F:29-42/Domain: amino-terminal nonhelical, NH1 <NHT>
F:43-1438/Region: interrupted helical
F:791-793/Region: cell attachment (R-G-D) motif
F:966-998/Region: cell attachment (R-G-D) motif
F:1154-1156/Region: cell attachment (R-G-D) motif
F:1306-1308/Region: cell attachment (R-G-D) motif
F:1345-1347/Region: cell attachment (R-G-D) motif
F:1432-1434/Region: cell attachment (R-G-D) motif
F:1493-1670/Domain: carboxyl-terminal nonhelical, NC1 <NCT>
F:1491-1651/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:31,33,39,41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status predi
F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F:1505-1511,1616-1622/Disulfide bonds: #status predicted
F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Alignment Scores:
Pred. No.: 30.8 Length: 1670
Score: 56.50 Matches: 16
Percent Similarity: 54.05% Conservative: 4
Best Local Similarity: 43.24% Mismatches: 12
Query Match: 27.16% Indels: 5
DB: 1 Gaps: 3

US-09-513-999C-3792_COPY_51_161 (1-111) x CGHUB (1-1670)
QY 3 GGGTGATCTTTTGGCTTCAGAGATCTTTTCATCTTT-----GCAGGACTTCT 53
|||||
DB 1573 GATPTLaserleuTrpLysglPheSerPheIleMetPheThrSerIadIysGrcIu 1592
|||||

QY 54 GGGGGCCGA---GTATGTAACCTCGCTGGCTCTGTCGTGAGAGAGG 101
|||||
DB 1593 GATTTGlyGlnAlaLeuAlaSerProGlySer---CysIleuGlnIuPhe 1608
|||||

```

```
Search completed: April 30, 2003, 14:36:50
Job time : 17.5 secs
```

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 30, 2003, 14:34:50 ; Search time 32.5 Seconds

(without alignments)
910.204 Million cell updates/sec

Title: US-09-513-999c-3792_COPY_51_161

Perfect score: 208 1 atgggttgatcttttgctt.....gcttgatgctgctctact 111

Sequence:

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2_1/uspro.spool/us0951399/runal_30042003_143400_25586/app_query.fasta_1.263
-DB-A-Geneseq_101002 -OPMT=fastan -SUFFIX=n2p.rag -MIMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosun62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pco -NOM=ext -HAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0951399_@CGN_1.1.114.@runal_30042003_143400_25586 -NCPV=6 -ICPV=3
-NO_XLPMAP -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A.Geneseq_101002: *
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT: *
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT: *
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT: *
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT: *
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT: *
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT: *
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT: *
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT: *
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT: *
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT: *
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT: *
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT: *
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT: *
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT: *
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT: *
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT: *
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT: *
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT: *
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT: *
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT: *
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT: *
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT: *
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|-------|------------------------------|
| 1 | 193 | 92.8 | 131 | 21 | AA603788 Human secreted pro |
| 2 | 91 | 45.0 | 280 | 22 | ABG15067 Novel human diagno |
| 3 | 85 | 40.9 | 57 | 23 | ABP09726 Human ORFX protein |
| 4 | 80 | 39.6 | 352 | 22 | ABG27640 Novel human diagno |
| 5 | 78 | 37.5 | 93 | 22 | AAW24321 Human EST encoded |
| 6 | 67 | 33.2 | 864 | 22 | ABG07580 Novel human diagno |
| 7 | 67 | 33.2 | 864 | 22 | AAU32809 Novel human secret |
| 8 | 63.5 | 31.4 | 386 | 22 | AAU19510 Human extracellular |
| 9 | 63.5 | 31.4 | 640 | 20 | AAV29918 Human CHD1 protein |
| 10 | 63.5 | 31.4 | 648 | 20 | AAV29917 Human protein sequ |
| 11 | 63.5 | 31.4 | 648 | 22 | ABG95518 Novel human diagno |
| 12 | 63.5 | 31.4 | 1176 | 22 | ABG22576 Human EST encoded |
| 13 | 61 | 29.3 | 52 | 22 | AAW24448 Human extracellular |
| 14 | 60.5 | 29.1 | 77 | 22 | AAW48057 Partial sequence o |
| 15 | 60.5 | 29.1 | 218 | 16 | AAW79164 Human type IV coll |
| 16 | 60.5 | 29.1 | 218 | 20 | AAV44172 Human alpha3 type |
| 17 | 60.5 | 29.1 | 218 | 21 | AAV56784 Human alpha-3 chn1 |
| 18 | 60.5 | 29.1 | 218 | 22 | AAE09484 Type IV collagen N |
| 19 | 60.5 | 29.1 | 268 | 20 | AAV31993 Human alpha3(IV)N |
| 20 | 60.5 | 29.1 | 268 | 21 | AAV97555 Partial sequence o |
| 21 | 60.5 | 29.1 | 471 | 16 | AAW79163 Bovine type IV col |
| 22 | 60.5 | 29.1 | 471 | 20 | AAV44171 Bovine alpha3 type |
| 23 | 60.5 | 29.1 | 471 | 21 | AAV56783 Bovine alpha-3 cha |
| 24 | 60.5 | 29.1 | 471 | 22 | AAE09483 Herbidically activ |
| 25 | 60 | 28.8 | 1145 | 23 | ABG95513 Novel human diagno |
| 26 | 59 | 29.2 | 854 | 22 | ABG28918 Mammalian son of s |
| 27 | 59 | 29.2 | 1297 | 14 | AAW47048 MSOS2 protein. Mu |
| 28 | 59 | 29.2 | 1297 | 16 | AAW84639 C glutathione prote |
| 29 | 58.5 | 29.0 | 126 | 22 | AAW92343 Drosophila melanog |
| 30 | 58.5 | 29.0 | 583 | 22 | ABW71010 Human epidermal pr |
| 31 | 58.5 | 29.0 | 668 | 20 | AAV32029 Macaca mulatta rha |
| 32 | 58 | 27.9 | 299 | 21 | AAW53137 Novel human diagno |
| 33 | 58 | 27.9 | 887 | 22 | ABG23684 Drosophila melanog |
| 34 | 58 | 28.7 | 1099 | 22 | ABW68903 Novel human diagno |
| 35 | 58 | 27.9 | 1113 | 22 | ABG06702 Human secreted pro |
| 36 | 57 | 27.4 | 61 | 22 | AAU01610 Gene #12 associate |
| 37 | 57 | 28.2 | 505 | 22 | AAW80430 Secreted protein e |
| 38 | 57 | 28.2 | 509 | 22 | AAW62173 Secreted protein e |
| 39 | 57 | 28.2 | 509 | 22 | AAW80382 Human albumin fusi |
| 40 | 57 | 28.2 | 509 | 23 | ABG63574 Human albumin fusi |
| 41 | 57 | 28.2 | 554 | 23 | ABG65272 Secreted protein e |
| 42 | 57 | 28.2 | 554 | 22 | AAW80407 Human albumin fusi |
| 43 | 57 | 28.2 | 554 | 23 | ABG65271 Human polypeptide, |
| 44 | 57 | 28.2 | 594 | 22 | AAW93619 Human normal bladd |
| 45 | 57 | 28.2 | 783 | 20 | AAV60344 |

ALIGNMENTS

RESULT 1
ID AAG03788 standard; Protein; 131 AA.
XX AAG03788;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein, SEQ ID NO: 7869.
XX
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
XX 06-SEP-2000.
XX

PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI: 2000-500381/45.
XX N-PSDB: AAC03794.
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS Claim 13; SEQ ID 7869; 71pp + CD-ROM; English.
XX
CC The present sequence is a polypeptide encoded by one of a large number
CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs
CC were prepared from total human RNAs or polyA+ RNAs derived from 30
CC different tissues. EST sequences usually correspond mainly to the 3'
CC untranslated region (UTR) of the mRNA because they are often obtained
CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for
CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in
CC those cases where longer cDNA sequences have been obtained, the full 5'
CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'
CC ends and can therefore be used to obtain full length cDNAs and genomic
CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and
CC chromosome mapping procedures. They are used to obtain upstream
CC regulatory sequences and to design expression and secretion vectors.
XX
SQ Sequence 131 AA:

Alignment Scores:
Pred. No.: 1-58e-16 Length: 131
Score: 193.00 Matches: 37
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 92.79% Indels: 0
DB: 21 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x AAG03788 (1-131)
QY 1 ATGGTGCATCTTTTGCCTTCAGAGATCTTTTCATCTTTCAGAGGACTTGGGGCCG 60
DB 1 MetGlyGlySerPheAlaLeuGlnAspSerPheSerSerLeuGlnGlyLeuGlyPro 20
QY 61 GAGTATGTAAGACCTGGGTCTGTGTGCTGCTGAGTGGCTGCTCTACT 111
DB 21 GluTyValLysLeuLeuGlyLeuGlyValCysLeuSerGlyCysSerThr 37

RESULT 2
ABG15067
ID ABG15067 standard; Protein: 280 AA.
XX
XX ABG15067;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #15058.
XX
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639363/73.
XX N-PSDB: AAS79234.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
PS Claim 20; SEQ ID NO 45426; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations in
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 280 AA:

Alignment Scores:
Pred. No.: 0.00247 Length: 280
Score: 91.00 Matches: 17
Percent Similarity: 63.64% Conservative: 4
Best Local Similarity: 51.52% Mismatches: 12
Query Match: 45.05% Indels: 0
DB: 22 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x ABG15067 (1-280)
QY 100 CACTCAGGCACACACAGACAGAGATTTTACATCTCCGCCCCAGAACTCCCTGCA 41
DB 23 HisSerAspThrTyArgTyrLeuGlyAlaLeuAspThrProAlaProValPheProThr 42
QY 40 AAGATGAAAAAGATCTTCAGAGGCAAGATCCACCA 2
DB 43 LysValSerAlaThrProAlaLysGlnLysValAlaGPro 55

RESULT 3
ABP09726
ID ABP09726 standard; Protein: 57 AA.
XX
XX ABP09726;
XX
XX 24-JUN-2002 (first entry)
XX
XX Human ORFX protein sequence SEQ ID NO:19434.
XX
XX Human: open reading frame; ORFX; gene therapy; cancer; cirrhosis;
XX hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
XX degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX hypertension; hypothyroidism; cholesterol ester storage disease;
XX

| | |
|--|--|
| XX | ABG27640; |
| AC | |
| XX | |
| DT | 18-FEB-2002 (first entry) |
| XX | |
| DE | Novel human diagnostic protein #27631. |
| XX | |
| KW | Human; chromosome mapping; gene mapping; gene therapy; forensic; |
| KN | food supplement; medical imaging; diagnostic; genetic disorder. |
| OS | Homo sapiens. |
| PV | WO200175067-A2. |
| XX | |
| PD | 11-OCT-2001. |
| XX | |
| PF | 30-MAR-2001; 2001WO-US08631. |
| PR | 31-MAR-2000; 2000US-0540217. |
| PR | 23-AUG-2000; 2000US-0649167. |
| PA | (HYSE-) HXSEQ INC. |
| XX | |
| PI | Dymanac RT, Liu C, Tang YT; |
| XX | |
| DR | WPI; 2001-639362/73. |
| DR | N-PSDB: MAS91827. |
| PT | New isolated polynucleotide and encoded polypeptides, useful in |
| PT | diagnostics, forensics, gene mapping, identification of mutations |
| PT | responsible for genetic disorders or other traits and to assess |
| PT | biodiversity - |
| XX | |
| PS | Claim 20; SEQ ID No 57999; 103bp; English. |
| XX | |
| CC | The invention relates to isolated polynucleotide (I) and |
| CC | polypeptide (II) sequences. (I) is useful as hybridisation probes, |
| CC | polymerase chain reaction (PCR) primers, oligomers, and for chromosome |
| CC | and gene mapping, and in recombinant production of (II). The |
| CC | polynucleotides are also used in diagnostics as expressed sequence tags |
| CC | for identifying expressed genes. (I) is useful in gene therapy techniques |
| CC | to restore normal activity of (II) or to treat disease states involving |
| CC | ((II)). (II) is useful for generating antibodies against it, detecting or |
| CC | quantitating a polypeptide in tissue, as molecular weight markers and as |
| CC | a food supplement. (II) and its binding partners are useful in medical |
| CC | imaging of sites expressing (II). (I) and (II) are useful for treating |
| CC | disorders involving aberrant protein expression or biological activity. |
| CC | The polypeptide and polynucleotide sequences have applications in |
| CC | diagnostics, forensics, gene mapping, identification of mutations |
| CC | responsible for genetic disorders or other traits to assess biodiversity |
| CC | and to produce other types of data and products dependent on DNA and |
| CC | amino acid sequences. ABG00010-ABG30377 represent novel human |
| CC | diagnostic amino acid sequences of the invention. |
| CC | Note: The sequence data for this patent did not appear in the printed |
| CC | specification, but was obtained in electronic format directly from WIPO |
| CC | at ftp.wipo.int/pub/published_pct_sequences. |
| XX | |
| SQ | Sequence 352 AA: |
| | |
| Alignment Scores: | |
| Pred. No.: | 0.0661 Length: 352 |
| Score: | 80.00 Matches: 16 |
| Percent Similarity: | 62.50% Conservative: 4 |
| Best Local Similarity: | 50.00% Mismatches: 12 |
| Query Match: | 39.60% Indels: 0 |
| DB: | Gaps: 0 |
| US-09-513-999C-3792_COPY_51_161 (1-111) x ABG27640 (1-352) | |
| XY | 106 AGCAGCCACTGAGGCACACAGAGACCGAGGTTCATCACTAATGGGCCACAGAAAGTCTC 47 |
| db | 190 SerserYusssergiYmethinistylsAsrpleUglAlaleuchnietThrProalaoPrcOlsyer 209 |

```

QY 46 CCTGCAAGATGAAAAAGATCTCTGCAGGCAAG 11
DB 210 ProthrAsnValAlaAlaIthrGlnAlaArgArgGlu 221
RESULT 5
ID AAM24321 standard; Protein; 93 AA.
AC AAM24321;
XX
XX 12-OCT-2001 (first entry)
DE Human EST encoded protein SEQ ID NO: 1846.
XX
XX Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KM diagnostics; forensic test; gene mapping; genetic disorder;
XX biodiversity; gene therapy; nutrition.
XX Homo sapiens.
XX
XX WO200154477-A2.
XX
XX 02-AUG-2001.
XX
XX 25-JAN-2001; 2001MO-US02687.
XX
XX 25-JAN-2000; 2000US-0491404.
XX 17-JUL-2000; 2000US-0617746.
XX 03-AUG-2000; 2000US-0631451.
XX 15-SEP-2000; 2000US-0663870.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
XX Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
XX WPI: 2001-476164/51.
XX N-PSDB; AAB98980.
XX
XX Isolated polypeptide for treatment of diseases, diagnostics, raising
XX antibodies and research use -
XX
XX Claim 20; Page 1198; 1275pp; English.
XX
XX The present invention provides the protein and coding sequences of novel
XX proteins from a variety of organisms, including human, dog, cat, horse,
XX cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
XX urchin and tomato. These were derived from expressed sequence tags (ESTs)
XX from the organism of interest. They can be used in diagnostics,
XX forensics, gene mapping, identification of mutations, to assess
XX biodiversity and for nutritional purposes. The present sequence is a
XX protein of the invention.
XX
XX SQ Sequence 93 AA:
XX
XX Alignment Scores:
XX Pred. No.: 0.108 Length: 93
XX Score: 78.00 Matches: 17
XX Percent Similarity: 71.43% Conservative: 3
XX Best Local Similarity: 60.71% Mismatches: 8
XX Query Match: 37.50% Indels: 0
XX DB: Gaps: 0
XX
US-09-513-999c-3792_COPY_51_161 (1-111) x AAM24321 (1-93)
QY 16 GCCTTGACAGATTCCTTTTCATTCCTTGCAGGACTCTGGGCGCGAGATGTAAGATC 75
DB 62 AlaIeuProGluMetProSerProPheIeuClyIleIeuArgIeuGluIuTyValLysLeu 81
QY 76 CTGGGTCTCTGTGTGTGCTGAGT 99
DB 82 LeuGlyIeuCysMetCysIeuSer 89

```

```

RESULT 6
ID ABG07580 standard; Protein; 864 AA.
XX
XX ABG07580;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #7571.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001MO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
XX N-PSDB; AAS71767.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity -
XX
XX Claim 20; SEQ ID NO 37939; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG030377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX atftp.wipo.int/pub/published_pcl_sequences.
XX
XX SQ Sequence 864 AA:
XX
XX Alignment Scores:
XX Pred. No.: 3.37 Length: 864
XX Score: 67.00 Matches: 16
XX Percent Similarity: 58.06% Conservative: 2
XX Best Local Similarity: 51.61% Mismatches: 13
XX Query Match: 33.17% Indels: 0
XX DB: Gaps: 0
XX
US-09-513-999c-3792_COPY_51_161 (1-111) x ABG07580 (1-864)

```

OY 98 CTCAGGCACACAGAGCCAGAGTTTACATCTCCGCCCCAGAGTCCCTGCAGAA 39
 ||||| ::||| |||||
 Db 696 LeuArgGlyProProGlnLeuGlnSerGlnLeuValSerGlyProArgSerProLysSer 715
 OY 38 GATGAAAAAGATCCTGCAGGCAAAAGATCCA 6
 ||| ||||| ||| |||||
 Db 716 AspSerGlyGlnSerCysLeuAlaTrpAspPro 726

RESULT 7
 AAU32809
 ID AAU32809 standard; Protein: 864 AA.
 AC AAU32809;
 XX
 XX 18-DEC-2001 (first entry)
 DE Novel human secreted protein #3300.
 XX
 XX Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX Homo sapiens.
 OS
 XX WO200179449-A2.
 XX
 XX 25-OCT-2001.
 PD
 XX 16-APR-2001; 2001WO-US08656.
 PF
 XX 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX Tang YT, Liu C, Drmanac RT;
 PI
 XX WPI; 2001-611725/70.
 DR
 XX Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 PT
 XX Claim 20; Page 668; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukemias. AAU29510-AAU3304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 CC
 XX
 SO Sequence 864 AA;

Alignment Scores:
 Pred. No.: 3.37 Length: 864
 Score: 67.00 Matches: 16
 Percent Similarity: 58.06% Conservative: 2
 Best Local Similarity: 51.61% Mismatches: 13
 Query Match: 33.17% Indels: 0
 DB: 22 Gaps: 0

US-09-513-999c-3792_COPY_51_161 (1-111) x AAU32809 (1-864)

OY 98 CTCAGGCACACAGAGCCAGAGTTTACATCTCCGCCCCAGAGTCCCTGCAGAA 39
 ||||| ::||| |||||
 Db 696 LeuArgGlyProProGlnLeuGlnSerGlnLeuValSerGlyProArgSerProLysSer 715
 OY 38 GATGAAAAAGATCCTGCAGGCAAAAGATCCA 6
 ||| ||||| ||| |||||
 Db 716 AspSerGlyGlnSerCysLeuAlaTrpAspPro 726

RESULT 8
 AAU19510
 ID AAU19510 standard; Protein: 386 AA.
 AC AAU19510;
 XX
 XX 04-DEC-2001 (first entry)
 DE
 XX Human diagnostic and therapeutic polypeptide (DITHP) #96.
 DE
 XX Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
 KW cell proliferative disorder; Crohn's disease; lymphoma; leukaemia;
 KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
 KW respiratory disorder.
 KW
 XX Homo sapiens.
 OS
 XX WO200162927-A2.
 XX
 XX 30-AUG-2001.
 PD
 XX 21-FEB-2001; 2001WO-US06059.
 PF
 XX 24-FEB-2000; 2000US-0184693.
 PR 24-FEB-2000; 2000US-0184697.
 PR 24-FEB-2000; 2000US-0184698.
 PR 24-FEB-2000; 2000US-0184768.
 PR 24-FEB-2000; 2000US-0184769.
 PR 24-FEB-2000; 2000US-0184770.
 PR 24-FEB-2000; 2000US-0184771.
 PR 24-FEB-2000; 2000US-0184772.
 PR 24-FEB-2000; 2000US-0184773.
 PR 24-FEB-2000; 2000US-0184774.
 PR 24-FEB-2000; 2000US-0184776.
 PR 24-FEB-2000; 2000US-0184777.
 PR 24-FEB-2000; 2000US-0184797.
 PR 24-FEB-2000; 2000US-0184813.
 PR 24-FEB-2000; 2000US-0184837.
 PR 24-FEB-2000; 2000US-0184841.
 PR 24-FEB-2000; 2000US-0185213.
 PR 24-FEB-2000; 2000US-0185216.
 PR 12-MAY-2000; 2000US-0203785.
 PR 15-MAY-2000; 2000US-0204226.
 PR 16-MAY-2000; 2000US-0204525.
 PR 16-MAY-2000; 2000US-0204821.
 PR 16-MAY-2000; 2000US-0204908.
 PR 16-MAY-2000; 2000US-0205232.
 PR 17-MAY-2000; 2000US-0204815.
 PR 17-MAY-2000; 2000US-0204863.
 PR 17-MAY-2000; 2000US-0205221.
 PR 17-MAY-2000; 2000US-0205285.
 PR 17-MAY-2000; 2000US-0205286.
 PR 17-MAY-2000; 2000US-0205287.
 PR 17-MAY-2000; 2000US-0205323.
 PR 17-MAY-2000; 2000US-0205324.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
 XX Chen A, D'Sa SA, Ameshey S, Dahl CR, Dam TC, Daniels SE;
 PI Dutour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;
 PI Liu TF, Roseberry AM, Rosen BH, Russo PD, Stockdrener TK, Daffo A;
 PI Wright RJ, Yap PE, Yu JT, Bradley DL, Bratcher SK, Chen W,
 PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson SJ;

XX WPI: 2001-502867/55.
DR N-PSDB: AAS31081.
XX Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics
PT
XX
PS Claim 27; Page 452-453; 522pp; English.
XX
CC The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DTRP) polypeptides (II), which include e.g. enzymes,
CC and proteins involved in growth and development and receptors. (I) and
CC (II) may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DTRP expression. For example, (I) and
CC (II) may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DTRPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II)
CC may be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DTRPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DTRPs and in assays to identify modulators of DTRP
CC expression and activity. The anti-DTRP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DTRP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DTRPs in samples (e.g. by enzyme linked immunosorbant
CC assay (ELISA)). AAU19415-AAU19625 represent human diagnostic and
CC therapeutic (DTRP) polypeptides of the invention.
XX
SQ Sequence 386 AA;
SO
Alignment Scores:
Pred. No.: 8.98 Length: 386
Score: 63.50 Matches: 16
Percent Similarity: 66.67% Conservative: 8
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 31.44% Indels: 5
DB: Gaps: 3
US-09-513-999c-3792_COPY_51_161 (1-111) x AAU19510 (1-386)
Y 110 GTAGAGCAGCCACTGACGACACAGAGAGCCAGG-----AGTTTACATCTCCGCG 57
Db 314 IIEGInGUPro--GInGUThrGInGUProGInGUleuSerPheThrTyThrGly 332
QY 56 CCCGAGAGTCCCTGCAAGATGAAAGAAATCTCGAAGCAAAAGAT 9
Db 333 AsparGSer-----LysAspGInGUInGUcysLeuGInGUInGUasp 346
RESULT 9
ID AAY29918 standard; Protein: 640 AA.
XX
AC AAY29918;
XX
DT 18-NOV-1999 (first entry)
XX
DE Human CHD1 protein encoded by transcript cDNA2.
XX
KW Human: coronary heart disease susceptibility gene; CHD1; mutation;
KW chromosome 11; diagnosis; screening; PCR primer; metabolic disorder;
KW detection; hypolipidoproteinaemia; familial combined hyperlipidaemia;
KW insulin resistant syndrome X; multiple metabolic disorder; obesity;
KW diabetes; dyslipidaemic hypertension.
XX
OS Homo sapiens.

XX
PN MO945112-A2.
XX
PD 10-SEP-1999.
XX
PF 04-MAR-1999; 99WO-US04682.
XX
PR 04-MAR-1998; 98US-0034941.
XX
PR 06-APR-1998; 98US-0080934.
XX
PA (MYRI) MYRIAD GENETICS INC.
XX
PI Ballinger DG, Ding W, Wagner S, Hess MA;
XX
DR WPI: 1999-540844/45.
DR N-PSDB: AA226833.
XX
PS Claim 10; Page 124-125; 297pp; English.
XX
CC The present invention describes the human chromosome 11-linked coronary
CC heart disease susceptibility gene (CHD1). Mutations in the CHD1 locus
CC in the germline are indicative of a predisposition to coronary heart
CC disease or to metabolic disorders related to lipid metabolism.
CC Products from the present invention can be used in the diagnosis
CC of predisposition to coronary heart disease and to metabolic disorders,
CC including hypolipidoproteinaemia, familial combined hyperlipidaemia,
CC insulin resistant syndrome X or multiple metabolic disorder, obesity,
CC diabetes and dyslipidaemic hypertension. CHD1 proteins can be used for
CC treating coronary heart disease and metabolic disorders. The products
CC can also be used for detection and drug screening. AA226832 to AA226841
CC and AA227027 to AA227029 represent human CHD1 nucleotide sequences.
CC AA229917 to AAY29926 represent human CHD1 proteins and protein sequences
CC used in the exemplification of the present invention. AA226842 to
CC AA226862 represent primers used in the identification of human CHD1;
CC AA226863 to AA227014 represent PCR primers used in the screening of
CC mutations in human CHD1; AA227015 to AA227026 represent oligonucleotides
CC used in the exemplification of the present invention.
XX
SQ Sequence 640 AA;
SO
Alignment Scores:
Pred. No.: 9.32 Length: 640
Score: 63.50 Matches: 16
Percent Similarity: 66.67% Conservative: 8
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 31.44% Indels: 5
DB: Gaps: 3
US-09-513-999c-3792_COPY_51_161 (1-111) x AAY29918 (1-640)
QY 110 GTAGAGCAGCCACTGACGACACAGAGAGCCAGG-----AGTTTACATCTCCGCG 57
Db 292 IIEGInGUPro--GInGUThrGInGUProGInGUleuSerPheThrTyThrGly 310
QY 56 CCCGAGAGTCCCTGCAAGATGAAAGAAATCTCGAAGCAAAAGAT 9
Db 311 AsparGSer-----LysAspGInGUInGUcysLeuGInGUInGUasp 324
RESULT 10
ID AAY29917 standard; Protein: 648 AA.
XX
AC AAY29917;
XX
DT 18-NOV-1999 (first entry)
XX
DE Human CHD1 protein encoded by transcript cDNA1.
XX
KW Human: coronary heart disease susceptibility gene; CHD1; mutation;
KW

XX chromosome 11; diagnosis; screening; PCR primer; metabolic disorder;
KM detection; hypocalphaipoproteinemia; familial combined hyperlipidaemia;
KM insulin resistant syndrome X; multiple metabolic disorder; obesity;
KM diabetes; dyslipidaemic hypertension.
XX
XX Homo sapiens.
XX OS
XX MO9945112-A2.
XX PN
XX 10-SEP-1999.
XX PD
XX PF 04-MAR-1999; 99WO-US04682.
XX PR 04-MAR-1998; 98US-0034941.
XX PR 06-APR-1998; 98US-0080934.
XX PA (MYRI-) MYRIAD GENETICS INC.
XX PI Ballinger DG, Ding W, Wagner S, Hess MA;
XX WP1: 1999-540844/45.
XX DR N-PSDB; AA226832.
XX JR
XX
XX New isolated coronary heart disease susceptibility gene, used to
PT develop products for diagnosis and treatment of coronary heart disease
PT and metabolic disorders -
XX
XX
PS Claim 11; Page 124; 297pp; English.
XX
XX The present invention describes the human chromosome 11-linked coronary
CC heart disease susceptibility gene (CHD1). Mutations in the CHD1 locus
CC in the genome are indicative of a predisposition to coronary heart
CC disease or to metabolic disorders related to lipid metabolism.
CC Products from the present invention can be used in the diagnosis
CC of predisposition to coronary heart disease and to metabolic disorders,
CC including hypocalphaipoproteinemia, familial combined hyperlipidaemia,
CC insulin resistant syndrome X or multiple metabolic disorder, obesity,
CC diabetes and dyslipidaemic hypertension. CHD1 proteins can be used for
CC treating coronary heart disease and metabolic disorders. The products
CC can also be used for detection and drug screening. AA226832 to AA226841
CC and AA227027 to AA227029 represent human CHD1 nucleotide sequences.
CC AA229917 to AA229926 represent human CHD1 proteins and protein sequences
CC used in the exemplification of the present invention. AA226842 to
CC AA226862 represent primers used in the identification of human CHD1;
CC AA226863 to AA227014 represent PCR primers used in the screening of
CC mutations in human CHD1; AA227015 to AA227026 represent oligonucleotides
CC used in the exemplification of the present invention.
XX
XX
XX Sequence 648 AA;
SQ
XX
XX Alignment Scores:
Pred. No.: 9.33 Length: 648
Score: 63.50 Matches: 16
Percent Similarity: 66.67% Conservative: 8
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 31.44% Indels: 5
DB: 20 Gaps: 3
US-09-513-999c-3792_COPY_51_161 (1-111) x AA229917 (1-648)
OY 110 GTAGAGCAGCCACTGACGACACAGACCAGG-----AGTTTACATCTCCGCG 57
DB 300 IIEGlnGluPro---GlnGluThrGlnGluProGlnIleLeuSerPheThrTyrThrly 318
OY 56 CCCAGAGTCCCTGCCAAAGATGAAAAAGAAATCTTCGCAAGCAAAAGAT 9
DB 319 AsparGser-----LysAspGlnGlnGlnGlnGlnGlnGlnGlnGln 332
RESULT 11
AAB95518
ID AAB95518 standard; Protein; 648 AA.
AC AAB95518;

XX
XX 26-JUN-2001 (first entry)
XX
XX Human protein sequence SEQ ID NO:18091.
XX DE
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX KM
XX Homo sapiens.
XX OS
XX EP1074617-A2.
XX PN
XX 07-FEB-2001.
XX PD
XX PF 28-JUL-2000; 2000EP-0116126.
XX PR 29-JUL-1999; 99JP-0248036.
XX PR 27-AUG-1999; 99JP-0300253.
XX PR 11-JAN-2000; 2000JP-0118776.
XX PR 02-MAY-2000; 2000JP-0183767.
XX PR 09-JUN-2000; 2000JP-0241899.
XX PA
XX (HELT-) HELIX RES INST.
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX DR WP1: 2001-318749/34.
XX
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX
PS Claim 8; SEQ ID 18091; 2537pp + CD ROM; English.
XX
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX
XX Sequence 648 AA;
SQ
XX
XX Alignment Scores:
Pred. No.: 9.33 Length: 648
Score: 63.50 Matches: 16
Percent Similarity: 66.67% Conservative: 8
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 31.44% Indels: 5
DB: 22 Gaps: 3
US-09-513-999c-3792_COPY_51_161 (1-111) x AAB95518 (1-648)
OY 110 GTAGAGCAGCCACTGACGACACAGACCAGG-----AGTTTACATCTCCGCG 57
DB 300 IIEGlnGluPro---GlnGluThrGlnGlnGlnGlnGlnGlnGlnGlnGln 318

QY 56 CCCAAGATCCCTGCAAGATGAAAGATCTCTCAAGCAAAAGAT 9
||||| : : : : :
Db 319 AsparGSer-----LysAspGluGluGluCysLeuGluGlnIuasp 332

RESULT 12
ABG22576
ID ABG22576 standard; Protein; 1176 AA.
XX
AC ABG22576;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #22567.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB: AAS86763.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20: SEQ ID No 52935; 103bp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1176 AA;

Alignment Scores:
Pred. No.: 9.76 Length: 1176
Score: 63.50 Matches: 16
Percent Similarity: 66.67% Conservative: 8
Best Local Similarity: 44.44% Mismatches: 7
Query Match: 31.44% Indels: 5

DB: 22 Gaps: 3
US-09-513-999C-3792_COPY_51_161 (1-111) x ABG22576 (1-1176)

QY 110 GTAGAGACGCCCTAGGCACACAGAGACCCAGG-----AGTTTACTACTCCGCG 57
: : : : :
Db 352 lIeGInGluPro--GlnGluThrGInGluProGluIleLeuSerPheThrTyThrGly 370

QY 56 CCCAAGATCCCTGCAAGATGAAAGATCTCTCAAGCAAAAGAT 9
||||| : : : : :
Db 371 AsparGSer-----LysAspGluGluGluCysLeuGluGlnIuasp 384

RESULT 13
AAM24448
ID AAM24448 standard; Protein; 52 AA.
XX
AC AAM24448;
XX
DT 12-OCT-2001 (first entry)
XX
DE Human EST encoded protein SEQ ID NO: 1973.
XX
KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;
KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;
KW diagnostics; forensic test; gene mapping; genetic disorder;
KW biodiversity; gene therapy; nutrition.
XX
OS Homo sapiens.
XX
PN WO200154477-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02687.
XX
PR 25-JAN-2000; 2000US-0491404.
PR 17-JUL-2000; 2000US-0617746.
PR 03-AUG-2000; 2000US-0631451.
PR 15-SEP-2000; 2000US-0663870.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Qian XB, Wang Z, Chen R, Asundi V;
PI Cao Y, Drmanac RA, Zhang J, Werhman T;
XX
DR WPI: 2001-476164/51.
DR N-PSDB: AAH99107.
XX
PT Isolated polypeptide for treatment of diseases, diagnostics, raising
PT antibodies and research use -
XX
PS Claim 20: Page 1256; 1275bp; English.
XX
CC The present invention provides the protein and coding sequences of novel
CC proteins from a variety of organisms, including human, dog, cat, horse,
CC cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea
CC urchin and tomato. These were derived from expressed sequence tags (ESTs)
CC from the organism of interest. They can be used in diagnostics,
CC forensics, gene mapping, identification of mutations, to assess
CC biodiversity and for nutritional purposes. The present sequence is a
CC protein of the invention.
XX
SQ Sequence 52 AA;

Alignment Scores:
Pred. No.: 16.3 Length: 52
Score: 61.00 Matches: 13
Percent Similarity: 45.95% Conservative: 4
Best Local Similarity: 35.14% Mismatches: 14
Query Match: 29.33% Indels: 6
DB: 22 Gaps: 1

US-09-513-999C-3792_COPY_51_161 (1-111) x AAM24448 (1-52)

QY 9 ATCTTTTCCTTCGAGGATTC-----TTTTCATCTTTGAGGAGCT 50
 ::||| ||| ||||| ||| |||:::
 Db 6 ValhelaValLeuCySGIYPheLeuTyLeuCySPhenLeuPhePheSerVal 25
 QY 51 TCTGGGCGCGAGTATGTAACCTCGGTCCTGTGTGCTGCTGAGTGG 101
 ::||| ||||| ::||| ||| ||| |||
 Db 26 ThrGlnaGlyValSerGluProArgSerSerHisCysThrProAlaTrp 42
 RESULT 14
 AAB48057
 ID AAB48057 standard; protein; 77 AA.
 AC AAB48057;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 DE Human extracellular signaling molecule (EXCS) (ID 1288847CD1).
 XX
 KW Extracellular signaling molecule; EXCS; anti-inflammatory; human;
 KW immunosuppressive; cytosolic; neuroprotective; gastrointestinal;
 KW vitruicide; antibacterial; anti-HIV; human immunodeficiency virus;
 KW anti-infectility; cerebroprotective; nootropic; antiulcer; antifungal;
 KW anticonvulsant; tranquilizer; neuroleptic; vasotropic; gynecological;
 KW keratolytic; protozoacide; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WC200070049-A2.
 XX
 PD 23-NOV-2000.
 XX
 PF 19-MAY-2000; 2000MO-US13975.
 XX
 PR 19-MAY-1999; 99US-0134949.
 PR 15-JUL-1999; 99US-0144270.
 PR 30-JUL-1999; 99US-0146700.
 PR 04-OCT-1999; 99US-0157508.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Tang YT, Yue H, Lal P, Burford N, Bandman O, Baughn MR;
 PI Azimzal Y, Lu DAM, Patterson C;
 PI
 DR N-PSDB; AAC84293.
 DR
 XX
 PT New human extracellular signaling nucleic acids and polypeptides useful
 PT for diagnosing, treating and preventing infections and
 PT autoimmune/inflammatory disorders -
 XX
 PS Claim 1; Page 85; 114pp; English.
 XX
 CC The invention provides human extracellular signaling molecules (EXCS)
 CC and polynucleotides which identify and encode EXCS. EXCS can be
 CC expressed by standard recombinant methodology. The amino acid and nucleic
 CC acid sequences of EXCS are useful for diagnosing, treating and
 CC preventing infections and gastrointestinal (peptic ulcer, dysphagia,
 CC pancreatitis), neurological (e.g. epilepsy, ischemic cerebrovascular
 CC disease, stroke), reproductive (infertility, ovulatory defects,
 CC endometriosis), autoimmune/inflammatory (actinic keratosis, acquired
 CC immunodeficiency syndrome (AIDS), Addison's disease), and cell
 CC proliferative disorders including cancers (of the breast, adrenal gland,
 CC bone). They may also be used to treat fatal familial insomnia,
 CC nutritional and metabolic diseases of the nervous system, myopathies,
 CC mental disorders (anxiety, schizophrenia, mood), as well as infections
 CC caused by parasites (malaria, leishmania, trypanosoma), viral
 CC (adenovirus, coronavirus, flavivirus), bacterial (e.g. pneumococcus,
 CC staphylococcus, bacillus), and fungal (aspergillus, blastomycosis,
 CC dermatophytes) agents. The nucleic acids, polypeptides, antagonists,
 CC agonists, pharmaceutical compositions, and antibodies may also be used
 CC for treating or preventing disorders associated with increased or

CC decreased expression or activity of EXCS. EXCS polynucleotides may also
 CC be used to detect and quantify gene expression in biopsied tissues in
 CC which expression of EXCS may be correlated with the disease, to determine
 CC presence or excess expression of EXCS, to monitor regulation of EXCS
 CC levels during therapeutic intervention, to detect the presence of
 CC associated disorders, as targets in microarray, to generate hybridization
 CC probes, and to detect differences in gene sequences among normal, carrier
 CC or affected individuals. Antibodies may also be used in diagnosing
 CC disorders, in monitoring patients being treated with EXCS agonists,
 CC antagonists or inhibitors. Sequences AAB48057-B48082 represent the EXCS
 CC of the invention.
 XX
 SQ Sequence 77 AA;
 XX
 Alignment Scores:
 Pred. No.: 19.4 Length: 77
 Score: 60.50 Matches: 13
 Percent Similarity: 59.26% Conservatve: 3
 Best Local Similarity: 48.15% Mismatches: 10
 Query Match: 29.09% Indels: 1
 DB: 22 Gaps: 1
 US-09-513-999c-3792_copy_51_161 (1-111) x AAB48057 (1-77)
 QY 28 TCTTTTCATCTTCGAGGAGCTCTG---GGCGCGAGTATGTAACCTCTGCTCTC 84
 ::||| ||| ||||| ||| ||| |||:::
 Db 23 ThrPheLeuGlySerGlnGlyLeuLeuGlnSerProSerTyGluValGlyCys 42
 QY 85 TGTGTGTGCTGCTGAGTGGCTGC 105
 ||||| |||||
 Db 43 CysLeuMetThrArgGlyCys 49
 RESULT 15
 AAR79164
 ID AAR79164 standard; Protein; 218 AA.
 XX
 AC AAR79164;
 XX
 DT 22-DEC-1995 (first entry)
 XX
 DE Partial sequence of human alpha 3 chain of type IV collagen.
 XX
 KW Type IV collagen; alpha 3 chain; Alport syndrome; COL4A3 gene.
 XX
 OS Homo sapiens.
 XX
 PN US5424408-A.
 XX
 PD 13-JUN-1995.
 XX
 PF 30-NOV-1990; 90US-0621091.
 XX
 PR 30-NOV-1990; 90US-0621091.
 XX
 PA (UNITV) UNITV KANSAS MEDICAL CENT.
 PA (UYVA) UNITV YALE.
 XX
 PI Hudson BG, Morrison KE, Reeders ST;
 PI
 DR WPI; 1995-262631/34.
 DR N-PSDB; AAQ96291.
 XX
 CC cDNAs encoding human or bovine alpha-3 type 4 collagen peptide(s) -
 CC useful for detection and therapeutic removal of antibodies associated
 CC with Goodpasture syndrome
 XX
 PS Disclosure; Columns 7-10; 33pp; English.
 XX
 CC using the PCR with primers derived from each end of the known 27 AA
 CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp
 CC bovine genomic fragment was amplified. This fragment was then used
 CC to a bovine lens cDNA library and a 1.5 kb partial cDNA clone was
 CC obtd. (clone KMC15). This encodes 238 residues of the triple helical

CC collagenous domain and all 233 residues of the C-terminal non-
CC collagenous (NC1) domain of the alpha 3 (IV) chain. This bovine
CC cDNA clone was used to screen a human kidney cDNA library and a
CC 2.7 kb human cDNA clone (clone KMC27) was obtd. This clone encodes
CC 218 residues of the NC1 domain and a portion of the 3' UTR region
CC of the human alpha 3(IV) chain. The COL4A3 gene localises to
CC chromosome 2 and therefore mutations in COL4A3 cannot be
CC responsible for Alport syndrome which is X-linked. An isolated
CC and substantially pure nt. having the sequence in AAQ96291
CC is claimed.
CC
XX

SQ Sequence 218 AA;

Alignment Scores:

| | | | |
|------------------------|--------|---------------|-----|
| Pred. No.: | 21 | Length: | 218 |
| Score: | 60.50 | Matches: | 17 |
| Percent Similarity: | 56.76% | Conservative: | 4 |
| Best Local Similarity: | 45.95% | Mismatches: | 11 |
| Query Match: | 29.09% | Indels: | 5 |
| B: | 16 | Gaps: | 3 |

US-09-513-999c-3792_COPY_51_161 (1-111) x AAR79164 (1-218)

| | | | |
|----|-----|--|-----|
| QY | 3 | GGGTGATCTTTGCTTGCGAGATCTTTTCATCTT-----GCAGGGACTTCY | 53 |
| | | | |
| Db | 121 | glytrpileserleutplysglypheserPheleketpherhserAlaaglyserGlu | 140 |
| | | | |
| QY | 54 | GGGGCCGGA---GTATGTAACACTCCTGGGTCTCTGTGTGCGCTGAGTGG | 101 |
| | | | |
| Db | 141 | GlyAlaGlyAlaAlaLeuAlaSerProGlySer---CysLeuGluGluPhe | 156 |
| | | ::: | ::: |

Search completed: April 30, 2003, 14:38:40
Job time : 34.5 secs